Developing And Assessing Methods For Calcium Imaging Data

by

Joseph L. Catallini II

A thesis submitted to Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science

Baltimore, Maryland

April, 2020

Abstract

Calcium imaging is an exciting cutting-edge technology used to quantify neuronal activity across a set of several cells in a local field. In this thesis, we analyzed and developed methods for both the processing and analysis of *in vitro* calcium imaging from neuronal cells. The calcium imaging measures activity in a number of cells from patients with Pitt-Hopkins syndrome (PTHS) and control subjects. We proposed a cell segmentation procedure to determine the locations of the cells in the image and processing steps to extract traces from the time series of calcium imaging under different conditions, including those with introduction pharmacological drugs to change activity patterns. We then utilized previously-published methods for time-series segmentation, employed with a variety of parameters, to determine times of activity and compare it to manual segmentation. We recommend a semi-automated processing stream for cell segmentation based on thresholding a median-centered, standard deviation-scaled image. We recommend using a motif-based correlation method for temporal peak detection, as models using results from this method most often agree with the statistical inferences drawn from models fit on a fully manual pipeline, but note some caveats and discrepancies compared to manual methods.

Thesis Committee

Primary Readers

John Muschelli III (Primary Advisor)
Assistant Scientist
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Andrew E. Jaffe
Associate Professor
The Lieber Institute for Brain Development, and
Department of Mental Health
Johns Hopkins Bloomberg School of Public Health

Acknowledgments

I thank thesis advisor John Muschelli III, Ph.D., and thesis reader Andrew Jaffe, Ph.D., for helpful conversations and advice on statistical analyses and the construction of this paper. I would like to thank the Lieber Institute for Brain Development (LIBD) for providing data and support in this study. In particular, I would like to pay special regards to Brady Maher, Ph.D., for collecting the calcium imaging data and providing support on the biological protocols of the experiment. I thank Stephanie Page, Ph.D. and Brittany Davis, Ph.D. for biological insight and assistance with manual selection. I wish to thank Madhavi Tippani, M.S., for extensive work and knowledge in calcium imaging processing and analysis at LIBD. I also wish to recognize the Biostatistics department at the Johns Hopkins Bloomberg School of Public Health for their formative influence in development of my statistical knowledge and practical application of said knowledge. So many professors and students from the department provided crucial support to this project and my pursuit of a Master's degree. Finally, I wish to express my deepest gratitude to my family and friends. I am especially grateful to my parents Lou and Alice, my sister Mary, and my partner Sarah. Thank you and I couldn't have done it without you!

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Measuring neuronal activity can give us insights into function and dysfunction of the brain. Calcium ions move in response to action potentials in neurons, which is the basis for the cutting-edge method of calcium imaging that records measurements of fluorescent intensity from high-resolution digital images as a proxy of neuronal activity (Grienberger and Konnerth, 2012). Applications of calcium imaging include measurements from *in vitro* and *in vivo* samples of neurons from human subjects and animal (primarily mice) samples, respecitvely. A major benefit of calcium imaging is that it allows for simultaneous measurements of large sets of neurons from the same subject in the same local field, as opposed to other methods for measuring neuronal activity, such as whole-cell electrophysiology, where typically a single neuron is measured at a time (Ahrens et al., 2013). However, these raw data require extensive processing to extract useful biological metrics of neuronal activity, and optimal parameter choices for these processing steps are under-explored. Therefore, in this project, we assessed the impacts of different processing steps on downstream statistical inference of neuronal activity using a set of calcium

imaging data from human induced pluripotent stem cell (iPSC)-derived and H9 embryonic stem cell (ESC)-derived neurons. These *in vitro* neuronal cell lines were further serially treated with different pharmacological agents to induce or block neuronal activity, which can serve as "ground truth" states to assess statistical bias in our ability to process and analyze calcium imaging data.

The surrogate measure of activity, fluorescence, in calcium imaging is given by light excitation of fluorescent proteins, or fluorophores, that have infected the cells and nuclei of neurons via a virus. Different fluorophores can identify specific areas of the cell to provide fluorescent signatures and can fluoresce at different wavelengths of light. Therefore, we can differentiate these fluorophores by applying different wavelengths of light excitation to the samples. Some fluorophores are designed to fluoresce more brightly in cell bodies at any state, and are thus effective in marking neuronal cells as "regions of interest" (ROIs). Other fluorophores differentiate calcium ion concentration by fluorescence, and are thus effective for measuring relative calcium ion concentration in a single ROI over time. Combining data from these two fluorophores allows us measure activity events and patterns over a time series in a single neuron/ROI. However, fluorescent intensity or fluorophore expression may differ across neurons for a variety of biological factors, so measurements of fluorescent intensity is likely most useful as a relative measure within each neuron over an experiment. Therefore, for each neuron/ROI, we estimate a baseline concentration of calcium and identify activity events as time points where the estimated calcium concentration is elevated from this

baseline (Dana et al., 2016).

These calcium imaging data measured via fluorescence are collected as a series of images and videos taken under high-resolution microscopes. Different microscopes can produce images with differing spatial and temporal resolutions. Carl Zeiss Microscopy (Oberkochen, Germany) develops microscopes for calcium imaging data recordings. Olympus (Shinjuku City, Tokyo, Japan) and Nikon (Minato City, Tokyo, Japan) are also large manufacturers of similar microscopes. Carl Zeiss microscopes export images as Carl Zeiss Image Data Files or CZI.

With this data, we would like to understand neuronal activity levels under a number of conditions. Furthermore, pharmacological agonists and antagonists may be introduced to the cells during a calcium imaging session that can increase or decrease activity. Estimating differences under these treatments can estimate differential responses across groups and can be used to inform algorithmic developments in identifying changes if a response to a pharmacological introduction is known to change activity (Maher and LoTurco, 2012). While we expect population-wide increased or decreased activity when exposed to these treatments, patient-derived neurons with pathology compared to neurotypical controls and different cell types may react differently to pharmacological treatments. Also, increased or decreased activity within a sample may be an indicator of neurological function. Additionally, we may be interested in the temporal relationship between neurons from a single subject, such as the correlation of the signal traces or the synchronicity of activity events. In this study, we are primarily interested in modulations of

neuronal activity when introduced to pharmacological conditions.

Identifying differences between patients with different neurodevelopmental or neuropsychiatric disorders and neurotypical controls is a common application of calcium imaging data (Mertens et al., 2015; Belinsky et al., 2014; Birey et al., 2017). The differences in activity between these phenotypes is likely smaller than the effects of different pharmacological treatements, further highlighting the need for robust data processing pipelines. For example, in the present study, some of these iPSC lines were derived from patients with Pitt-Hopkins syndrome (PTHS), an autism spectrum disorder that is characterized by intellectual disability and developmental delay, breathing problems, recurrent seizures (epilepsy), and distinctive facial features, and other lines were derived from neurotypical individuals. Pitt-Hopkins syndrome is very rare with approximately 500 reported cases worldwide, and is caused by mutations in the transcription factor 4 (TCF4) gene. There is no cure for PTHS, however treatment includes managing symptoms through medical and behavioral therapies (Sweetser et al., 2018; Amiel et al., 2007; Brockschmidt et al., 2007; Forrest et al., 2012; Sweatt, 2013; Van Balkom et al., 2012; Zweier et al., 2008).

Typical challenges with processing calcium imaging data - that are largely independent of biological contrasts being interrogated - include identifying regions of interest, movement correction and image alignment, selection and extraction of trace statistics, and identifying activity events (Giovannucci et al., 2019; Patel et al., 2015; Jewell et al., 2018). The primary aim of this thesis is to present and assess a generalizable and semi-automated pipeline for

processing calcium imaging data, identifying neurons, and identifying activity events in individual neurons. Multiple steps throughout this pipeline may require some input parameter. We will select and assess these parameters and methods utilized in this pipeline by their similarity to results generated from manual selection, with the assistance of a subject matter expert, and make a recommendation of methods and parameters to select for processing of future calcium imaging data sets. Finally, we will assess the processing pipeline by statistical inference generated from results and compared to statistical inference drawn from results of manual selection. We will present the thesis as a description of the methods and parameterizations employed in the pipeline, a summarization of the results of the pipeline compared to manual selection, a recommendation of the methods and parameterizations to employ in future data sets, and a discussion of the results and future work in calcium imaging.

All of this work is released as open-source software, which may use closed or proprietary software such as MATLAB and can be found in this Github repository. To help lower the burden on users, we also created a graphical user interface (GUI) with point-and-click abilities.

Chapter 2

Methods

2.1 Data

In this study, all of the data was collected from the Lieber Institute for Brain Development (LIBD, Baltimore, MD). We have data from 7 different cell lines derived from embryonic stem cells (N = 1), and human induced pluripotent stem cell (iPSC)-derived neurons from neurotypical control subjects (N = 3) and patients with Pitt-Hopkins syndrome (N = 3). Each cell line had 2 separate replicates. Neurons were cultured on glass coverslips, which were removed from culture media and continuously perfused with artificial cerebro-spinal fluid (ACSF) containing 128 mM NaCl, 30 mM glucose, 25 mM HEPES, 5 mM KCl, 2 mM $CaCl_2$, and 1 mM $MgCl_2$. The ACSF used in this study is described in Wen et al. (2014).

In each imaging session, the same local field was recorded. Throughout each session, up to 4 separate time series may have been recorded, each under a different pharmacologically-induced condition. In each session, the first condition was recorded with no pharmacological substances added to ACSF

perfusion described above. We refer to this condition as the baseline condition. In subsequent conditions, pharmacological substances were added to the ACSF perfusion. The first pharmacologically-induced condition introduced a $5\mu M$ solution of Gabazine to the field. Gabazine acts as an antagonist to gamma-Aminobutyric acid (GABA)-a receptors, which often inhibit neuronal activity. The introduction of Gabazine was intended to antagonize these inhibitory receptors, leading to an increase in activity. However, there have been reports of excitatory GABA receptors, which stimulate activity in certain cell populations, implying that introduction of Gabazine could lead to a decrease in activity in such populations (Taketo and Yoshioka, 2000; Cherubini, Gaiarsa, and Ben-Ari, 1991). The next pharmacological condition introduced a $10\mu M$ solution of 6,7-dinitroquinoxaline-2,3-dione (DNQX) mixed with a $100\mu M$ solution of (DL)-amino-5-phosphonovaleric acid (AP5) to the field. DNQX acts as an antagonist of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPAR) subtype of glutamate receptors. AP5 acts as an antagonist of the N-methyl-D-aspartate receptor (NMDAR) subtype of glutamate receptors. Together, these pharmacological agents should block all glutamatergic neurotransmission. In 3 sessions, a $1\mu M$ solution of a neurotoxin called tetrodotoxin (TTX) was introduced to the field. TTX is a potent antagonist of voltage-gated sodium channels and therefore blocks the generation of action potentials. This condition was intended to block all action potentials which are the major source of calcium events. These conditions allow us to determine how the event detection methods perform in condition of high, medium, and low-to-no activity. All pharmacological substances in this study are referenced in Maher and LoTurco (2012) and were purchased from Tocris

Cell Line	Group	DIV	Baseline	Gabazine	DNQXAP5	TTX
H9	ESC	66	2	1	1	0
H9	ESC	87	1	1	1	0
WT661	Control	66	1	0	1	0
WT661	Control	87	1	1	1	0
1001.5	Control	67	1	1	1	1
1001.5	Control	84	1	1	1	0
N11C2	Control	67	1	1	1	1
N11C2	Control	84	1	1	1	0
PH016	PTHS	66	1	1	1	0
PH016	PTHS	81	0	1	1	0
PH572	PTHS	67	2	1	1	1
PH572	PTHS	81	1	1	1	0
2005.1	PTHS	67	1	1	1	0
2005.1	PTHS	84	2	1	1	0
Total	–	–	17	14	14	3

Table 2.1: Description of cell line data. The first column provides a code for the donor. The second column provides the group of the donor (embryonic stem cell, control, or Pitt Hopkins). The third column provides the days *in vitro* (DIV) for the particular for that set of experiments. Columns 4 through 7 provide the number of fields successfully recorded in each condition. Note that each pharmacological condition should have a corresponding baseline condition taken from the same field. However, during the PH016 cell line at 81 DIV, a frame rate recording error during experimentation resulted in the baseline condition being removed from the data set.

Bioscience (Bristol, UK). For each of the 7 cell lines, 2 sessions of recordings were acquired, resulting in a total of 45 videos were recorded across 16 unique local fields of view (Table 2.1).

2.1.1 Acquisition and Preparation

The data were acquired using a custom-built Carl Zeiss Microscopy AxioExaminer Z.1 microscope equipped with Yokogawa (Musashino, Tokyo, Japan) Spinning Disk scanning unit, a Hamamatsu (Hamamatsu City, Japan) Flash4.0 V3 sCMOS camera and a 20X/1.0NA water immersion objective. Metadata

about the microscope, settings, and image, as well images themselves can be extracted using Bio-Formats, a software tool developed by the Open Microscopy Environment consortium and Linkert et al. (2010) that provides support to open source and commercial software programs. In this study, the data were exported into the Carl Zeiss Image (CZI) format. Metadata about the images frame rate, size, pixel units and channel information was extracted and images were converted to TIFF images using the Bioformats package in MATLAB. We converted to the TIFF format for interoperability and the ability to process images in a open source environment. If, in the future, we are to develop a fully open source pipeline, the conversion of data stored in a proprietary format, such as CZI, to an open source format will be a necessary first step.

In this study, the imaging data was recorded at a spatial resolution of $0.645 \times 0.645\,\mu m$ per pixel. The field of view for all images in the study was 640×1024 pixels, or $412.8 \times 660.48\,\mu m$.

During each imaging session, a reference image was acquired using fluorescence expressed from the mRuby protein. The mRuby protein, described in Kredel et al. (2009), is a cell-fill or cytosolic fluorophore expressed throughout the cytoplasm of the neuron. In this experiment, we used a neuronal-specific promoter to express the fluorophores only in neurons. This image was obtained under light excitation of wavelength 577 nm, referred to as the red channel. Under this light excitation, infected neurons were identified by mRuby fluorescence in the cell body. Areas where calcium was bound to the protein fluoresce brightly, indicating the presence of a neuron. However,

the mRuby fluorophore is not restricted to the cell body, so there may be areas of red channel images that fluoresce brightly, but do not identify the location of a neuron cell body, particularly neurite branches. Directly after this step, a dynamic video was obtained under light excitation of wavelength 493 nm, referred to as the green channel. Under this light excitation, the presence of the GCaMP6s protein is marked by fluorescence. The fluorescence in the image correlate with expression of the GCaMP6s protein where calcium ions are bound. Higher concentrations of calcium indicate elevated levels of GCaMP6s molecules in the fluorescent configuration, allowing this protein to dynamically report relative concentrations of calcium ions over time.

The imaging sessions in this study were organized as a pilot study and recorded across several days. The length of these videos range from 5 to 10 minutes for each condition and allow extraction of fluorescence over time from the regions of interest, or regions identified as neurons from the red channel. As the length of the recordings varied, we considered the length of each recording when drawing statistical inference of activity in neurons and how it relates to donor diagnosis and pharmacological condition. The fluorescence units are arbitrary across neurons, as different biological conditions of the neuron may impact the expression of the GCaMP6s protein. For this reason, we are interested in measuring the relative fluorescence in each neuron over time. Therefore, the absolute amplitude of the fluorescence peaks for each neuron were not of interest in this study. Instead, the primary focus was to identify neuronal activity events as timepoints when the fluorescence peaks occur (i.e. frequency of events). Videos recorded under the green channel were

recorded at 4 Hz. Throughout the image and video extraction, a wash of the ACSF described above flows over the cell regions. During the pharmacological intervention conditions, solutions of antagonist were introduced to the wash, with a wash-in period prior to beginning collection of images to allow the drug concentration in the bath to ready a steady-state equilibrium. All sets of cell lines were differentiated from iPSCs, as described in Wen et al. (2014), through a neural precursor cell (NPC) intermediary state into neurons across a time span of between 60 and 90 days *in vitro* prior to imaging data acquisition.

2.2 Processing Methods

2.2.1 Cell Segmentation

To identify the cells of the image, we used the red channel and defined ROIs by various semi-automated thresholding techniques. We considered a thresholding technique with manual intervention to be the true estimates of the cells (or at least the best estimate of the truth). We then applied various semi-automated thresholding techniques to assess which cell segmentation processing techniques most closely identify these true positions of the cells across images and experimental conditions.

2.2.1.1 Manual Segmentation: Best Estimate

A region growing method was implemented that requests user input in selecting interior pixels of each cell. The process continues by computing the mean intensity of the region and then includes the neighbor of the region with the most similar intensity, which continues until some threshold of the

difference of most similar neighbor is exceeded. The application allows for the user to select a cell, delete the previous selection, and adjust the threshold. A subject matter expert used this method to manually segment the red channel images in the data set. The results of this segmentation were treated as the best estimate of the truth in evaluating subsequent methods.

In our application of this method, we took the $\log_{10}$ transformation of the red channel image, and displayed this image for the user. This transformation enhances cell presentation as the data are highly right-skewed. Another transformation of the resulting image was stored as one minus min/max normalized version of the $\log_{10}$ image. In this method, the user viewed and clicked the $\log_{10}$ transformed image, however the region growing algorithm was performed on the inverted min/max normalized image. The user selected a seed for the region growing of the cell near the interior of the cell, which should be among the local minima of the region of interest in the inverted min/max normalized image. The region-growing process then began by computing the mean value of the region (the value of the seed pixel on the first step). The process then collected the neighboring pixels to the region that are not already in the region and selected the pixel with the value nearest to the mean region value. If the distance of the value of the new pixel to the region mean was less than a specified threshold, the pixel was added to the region. This process iterated until all remaining neighboring pixels of the region had a value with distance from the region mean greater than a user-selected threshold between 0 and 1. If the user was dissatisfied with the cell specification, the user removed the previously-defined ROI by simply

clicking on it. The user could then manipulate the threshold until the program produced a subjectively-accurate segmentation of the cell. This process was repeated by the user until all cells in each image were satisfactorily identified as ROIs. This procedure resulted in a binary mask with pixels valued 1 where an ROI was identified and 0 elsewhere. After this step the individual ROIs were uniquely labeled using the MATLAB function `bwlabel` (Haralick and Shapiro, 1992) and the user saved each labeled mask for future processing.

There are several important considerations for interpreting results from manual segmentation of collected calcium imaging data. First, manual cell segmentation allows for subject matter experts (trained neuroscientists) to identify plausible ROIs from its size and shape properties, as well as its fluorescence signal. This further reflects the intrinsic biases in collecting calcium imaging data, which usually does not involve sampling random fields of view, but instead focusing on, and recording from, fields containing numerous and putatively active cells. Therefore, data collection and manual segmentation from the same researcher represents a form of "gold standard" data with two forms of validation (collecting the data and calling ROIs) for evaluating ROI detection. For example, subject matter experts can identify cells that do not fluoresce brightly, as well as ignore regions that fluoresce brightly, but are likely imaging artifacts or other non-cellular bodies.

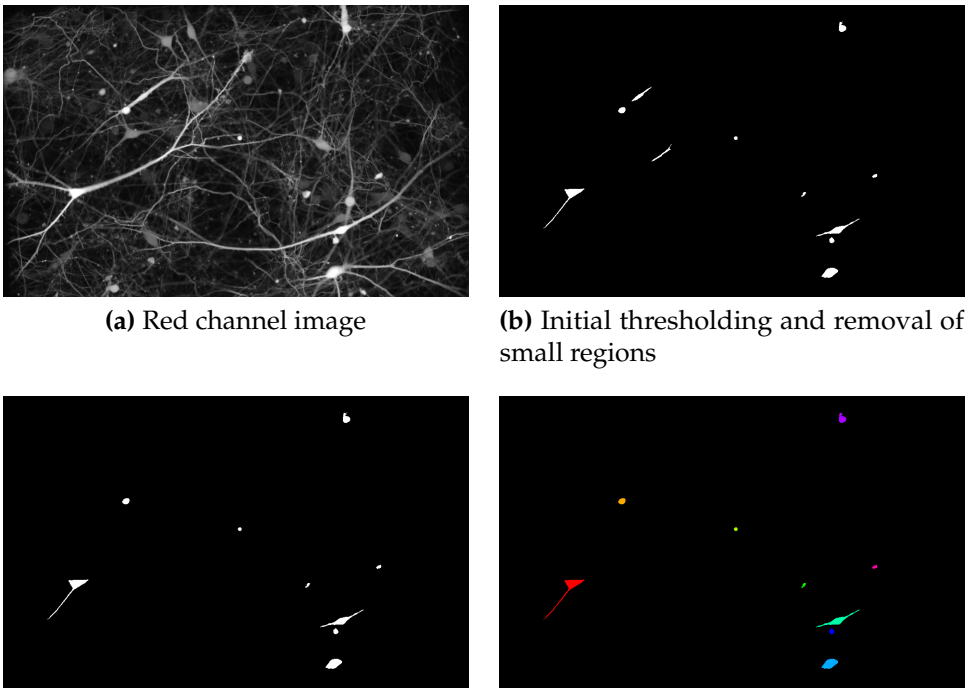
While there are numerous strengths of this approach, the most significant disadvantage to manual segmentation is that, in large data sets, this method requires significant time from the subject matter expert, which can quickly

become infeasible. Intra- and inter-reader variability also may cause differences in results, as results of manual segmentation may vary within a single reader on different days and may vary even moreso across different readers. Furthermore, these manual segmentations are often not collected in ways that permit combined analysis for differential activity across multiple or many images in an experiment. These weaknesses can be readily overcome with automated data processing approaches, but the absolute accuracy of such approaches has been largely understudied.

2.2.1.2 Semi-automated Cell Segmentation

Figure 2.1 displays a visualization of the steps of automated cell segmentation. In the figure, the method used is the median-eccentricity method.

Given the large numbers of cells and subsequent data generated in calcium imaging, the goal of this study was to develop and assess the accuracy of semi-automated pipelines for processing calcium imaging data. We therefore present semi-automated cell segmentation methods in two phases: an initial thresholding step, and a cell refinement step. Here we employed two procedures for the thresholding step and two procedures for the cell refinement step, resulting in four total cell segmentation methods. Both thresholding steps required a threshold parameter that takes as a multiplier of the standard deviation of the red channel image intensities as the input. We therefore examined the results from each method using a variety of thresholds to assess robustness to this important parameter by testing thresholds from 0.5 to 10, incrementing by 0.5. We then assessed each method and threshold by its



(a) Red channel image

(b) Initial thresholding and removal of small regions

(c) After watershed and removal of non-cellular regions

(d) Labeled mask

Figure 2.1: Visualization of automated cell segmentation. In panel (a) we see the $\log_{10}$ transformed red channel image. In actual processing, we use the raw image, however for visualization purposes, the $\log_{10}$ transform allows clearer visualization of the cells. In panel (b), we see a binary mask produced with small regions removed. In this example, the initial thresholding is done with the median-eccentricity method. In panel (c), we see the binary mask after a watershed algorithm is performed and regions that we do not believe are cellular bodies are removed. In this example, the eccentricity-based method is used to remove thes regions. In panel (d) we see the mask with each unique region of interest labeled and displayed by a unique color.

similarity to corresponding "gold standard" results from manual cell segmentation, and after this assessment, selected a single method and parameter for use throughout the remainder of the pipeline in this data set. Code for semi-automated cell segmentation is found in `segmentRed.m` in the .

2.2.1.2.1 Thresholding The first thresholding method involves first subtracting the red image median from the image and dividing by the image standard deviation; thus, pixel values represent standard deviations from the median. An initial binary mask was created by tresholding this image using a user-inputted threshold, S. Therefore, the initial mask was thresholded by S image standard deviations above the image median to produce potential ROIs. This step resulted in binary images valued 1 where putative cell bodies were potentially present, and valued 0 elsewhere.

The second thresholding method used the same region growing function as implemented in the manual segmentation method, but growing regions corresponding to background signal rather than cells. Here the user inputted a threshold, S, which was multiplied by the image standard deviation (σ) to produce the threshold inputted into the region growing method. Here, the background (pixels in the image where we do not believe there is a cell) was grown from a seed selected automatically (which is the inverse of the manual segmentation method above, where each cell region was grown from a seed selected by the user). In this method, the seed was selected among the pixels with minimum intensity values in the image. The region growing proceeded until all remaining neighboring pixels of the region had a value with distance from the region-mean greater than $S\sigma$. The result of this step is a binary image

valued 0 where putative cell bodies were potentially present and valued 1 elsewhere. This binary image is then inverted to mask ROIs (rather than the background).

2.2.1.2.2 Cell Refinement The above thresholding strategies provided two methods for generating ROI masks/locations which were based on a series of thresholds. We then combined these two thresholding methods with two additional refinement methods. One method was based on finding circles using the circular Hough transform (`imfindcircles` in MATLAB) (Atherton and Kerbyson, 1999; Yuen et al., 1990; Davies, 2005), as real cells are more likely to be circular objects in the binary masks. The other method was an eccentricity-based method, which estimates the conical shape of an area. The MATLAB function `regionprops` allows caculation of eccentricity of the bounding ellipse of each potential region of interest. The eccentricity of an ellipse is described by the following formula:

$$E = \frac{c}{a}$$

where E is the eccentricity of the ellipse, c is the distance from a focus point of the ellipse to the center of the ellipse, and a is the distance from that focus point to a vertex of the ellipse. A larger eccentricity implies that an ellipse is more oblong. The eccentricity of a perfect circle is 0, and the eccentricity of a straight line is 1.

Both cell refinement methods employed the MATLAB function `bwareaopen` to remove all clusters containing pixels below a pre-specified threshold, P. In this study, P was obtained using metadata extracted from the CZI file, which

included fields "ScaleX" and "ScaleY", or the length in µm of each dimension for a pixel in the image. We calculated P by the following equation:

$$P = \text{ceiling}(20/(S_X \times S_Y)) \tag{2.1}$$

where S_X is the length in µm of the x-dimension of a pixel and S_Y is the length in µm of the y-dimension of a pixel. This equation implies that in order to be considered a cell, the region of interest in the image must correspond to at least 20 μm^2 in area.

The cell refinement methods are described below.

1. Remove clusters with fewer than P pixels (2.1) using `bwareaopen`
2. If circle-finding method is employed, find circles in the resulting mask using `imfindcircles` with radius range $[\text{round}(\sqrt{P/\pi},\ 2 \times (\text{round}(\sqrt{P}) + 1)]$ where P is defined in 2.1. The lower bound of this value was calculated to be the radius of a circle with area in pixels equal to P. The upper bound of this value was selected by manual inspection of the red channel images to determine a reasonable upper bound for the radius of a cell in this study. After this range is estimated, it can be used for all images in the study. If the eccentricity-based method is employed, skip this step.
3. Perform watershed transformation (`watershed` in MATLAB) to separate overlapping cells as unique ROIs in the mask (Meyer, 1994).
4. If the eccentricity-based method is employed, use `regionprops` to calculate the eccentricity of the bounding ellipse of each potential ROI.

Exclude regions with eccentricity greater than 0.99. If the circle-finding method is employed, skip this step.

5. Use `bwareaopen` with threshold P once more to remove small clusters of pixels.

6. Use MATLAB function `imclearborder` with pixel connectivity 8 to remove potential ROIs near the border of the image (Soille, 2013).

7. Label the resulting binary mask using `bwlabel` to uniquely identify each ROI.

Throughout the rest of the paper, we refer to the method using the image transformed around the median and the eccentricity-based option as the "median-eccentricity method", the method using region growing and the eccentricity-based option as the "region growing-eccentricity method", the method using the image transformed around the median and the circle-finding option as the "median-circles method", and the method using region growing and the circle-finding option as the "region growing-circles method". Regardless of the method employed, this segmentation step results in ROIs that correspond to neuronal cells within each image, which were carried forward for downstream neuronal signal quantification.

In summary, we employed four total cell segmentation methods, each tested with 20 different threshold parameters, resulting in 80 total combinations. These combinations are illustrated in Figure 2.2.

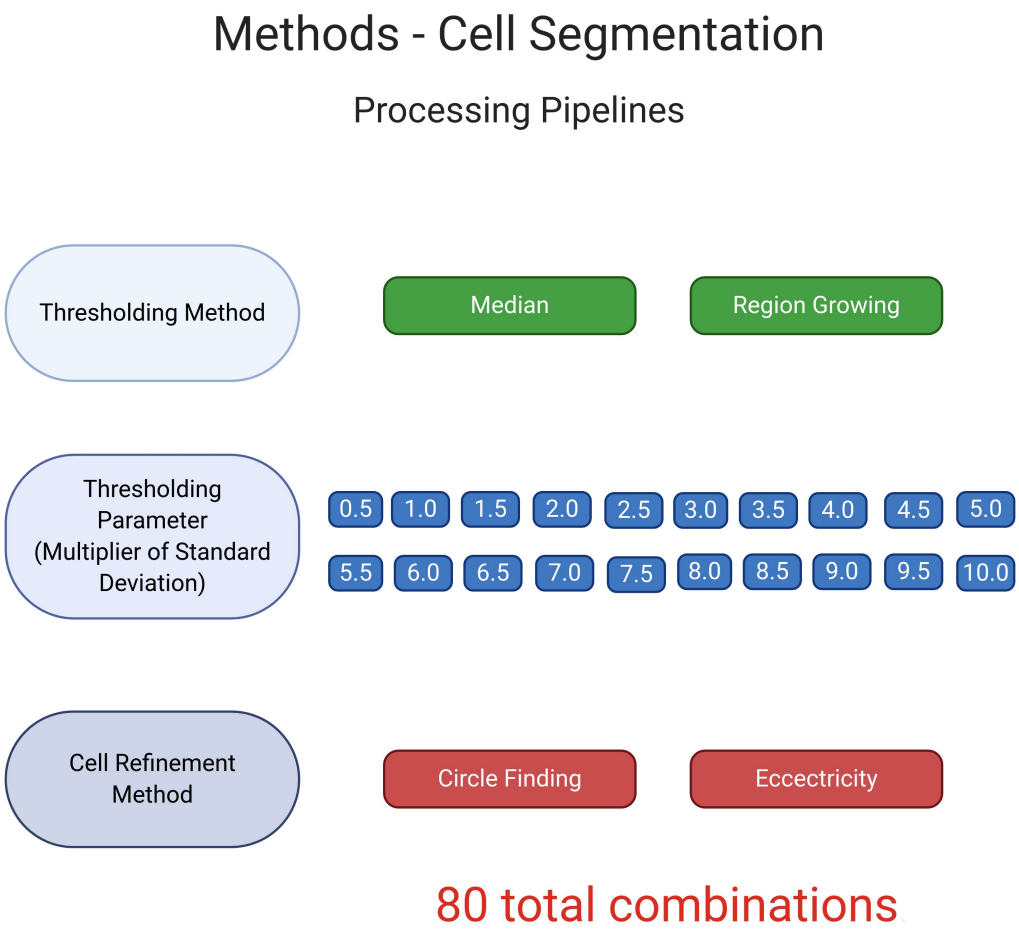


Figure 2.2: Visualization of all cell segmentation methods and parameterizations explored in this study.

2.2.2 Registration

After segmenting each cell/ROI in calcium imaging data, the next goal involved extracting the time series of its activity across the experiment. However, the fluorescence intensities (that correspond to activity) within each ROI may be noisy, as there may have been changes in the field of view during the activity recording, drift in the cells, and/or slight changes in the scope due to movements from the technician. Therefore, as ROIs can potentially move across the time series of activity measures, we performed image registration/alignment within each image to correct fluorescence tracings for these potential artifacts.

Within a time series experiment, multiple timepoints can serve as the reference for alignment. For these data, the first green channel (fluorescence activity) image has the least difference from the red channel (used to define ROIs), so we aligned all fluorescence/activity data to this first image for each ROI. All registrations included only a translation (and no rotations). We found that registering each image (time t) to the image at the time point before it (time $t-1$) was stable, as most changes are subtle or slow-moving. Let's assume the transformation from time t to $t-1$ is denoted as R_t, where $t = 2, \ldots T$, where T is the number of timepoints and no transformation is estimated for the first time point. As the transformation is a linear transformation, we can combine the transformations to provide a composed transformation to the first frame ($t = 1$). For example, for a time t, the combined transformation R_t^1 is

$$R_t^1 = R_t \times R_{t-1} \times \cdots \times R_2$$

where $\times$ is the matrix product. Using these composed matrices, each activity data set was registered to the first activity frame in that corresponding condition. In this step of the registration MATLAB's default monomodal registration configuration is used with the function `imregtform`. This configuration uses mean square error as a metric to estimate the registration matrix. After the green channel images are aligned, then the pixel-wise median over all time was calculated to generate a median frame.

Additionally, shifts in the scope field may occur between the acquisition of the red channel (for ROI detection) and the green channel (for fluorescence activity video). Moreover, multiple conditions may be recorded, increasing the probability the data are not perfectly aligned. Therefore, registering the activity images to the red channel is necessary when available. We estimated the registration matrix of the pixel-wise median image of the aligned green channel series to the corresponding red channel image using translation transformations. For the estimation of this registration matrix, we used MATLAB's default multimodal registration configuration in `imregtform`. This configuration uses Mattes mutual information (Mattes et al., 2003) as a metric to estimate the registration matrix. After the registration matrix was estimated, it was used to register each frame of the aligned green channel series to the red channel image. These registration procedures therefore produce normalized fluorescence intensity videos around each ROI. Code for time series registration is found in `registerGreen.m` in the CIRMIT github repository.

We evaluated the performance of the registration by visual inspection, and, for a fully-automated pipeline, estimated translation parameters can be

flagged. For example, if the estimated translation exceeds the approximate diameter of a cell, a potential error could be flagged. Additionally, the translation estimation can be used to determine if a given cell remained in the frame of the recording. In this study, for each identified cell, we recorded the length of time that the cell was visible (as defined by 1 pixel) in the frame of the recording. For example, if a cell went outside the field of view at time point 110, then the only time analyzed were points up to that time.

2.2.3 Trace Extraction

2.2.3.1 Summary Statistic

Once the activity data is registered and in the same space as the segmented red channel, we can simply use the masked ROIs to extract the time series for each ROI. To make a trace plot and identify activity events, we used a summary measure within each ROI at each time point. We note that the cell segmentations are not perfect, nor is the registration of the green to red channel, and thus there may be areas in the ROIs that are not cells, and are static (non-active) for the complete time series. We therefore explored different summary measures to identify activity, including the mean and quantiles of the ROI at each time point, including the 50^{th} (median), 75^{th}, and 95^{th} percentiles. Most commonly, the summary measure is the mean, but this measure is not resistant to outliers or small values induced by segmentation or registration errors. Code for calcium imaging trace extraction is found in `traceExtraction.m` in the CIRMIT github repository.

2.2.4 Smoothing: DFF

As protein expression and fluorescence vary across cells, we also employed a technique called "DFF", described in Jia et al. (2011) as $\frac{\Delta F}{F}$. This technique works on a fluoresence trace F and provides smoothing to the curves and can help with peak normalization. In this experiment, the DFF process began by computing a rolling five frame (1.25 second) average over each trace referred to as $\bar{F}$. The i^{th} component of $\bar{F}$ was computed as follows

$$\bar{F}_i = \text{mean}(\{F_k : k \in \mathbb{N}^{[i-2,i+2]\cap[1,T]}\})$$

where F_k was the k^{th} component of the raw trace and T was the total number of frames in the series. Next, a rolling minimum over a 251 frame (62.75 second) window of $\bar{F}$ was computed and referred to as F_0. The i^{th} component of F_0 is computed as follows

$$F_{0,i} = \min(\{\bar{F}_k : k \in \mathbb{N}^{[i-250,i]\cap[1,T]}\})$$

where $\bar{F}_k$ was the k^{th} component of $\bar{F}$. The i^{th} component of the DFF trace was computed as follows

$$DFF_i = \frac{\bar{F}_i - F_{0,i}}{F_{0,i}}$$

Code for DFF is found in `dff.m` in the CIRMIT github repository.

The windows sized used in this step were selected specifically for data recorded using the microscope in this study, however selecting window sizes

that correspond approximately to the real length of time of each (in seconds) for future data sets is recommended.

This DFF procedure therefore produced smoothed estimates of fluorescence activity within each ROI and video for subsequent activity characterization.

2.2.5 Event Detection

2.2.5.1 Manual Event Detection: Best Estimate

One of the most important steps of calcium imaging data processing is identifying discrete action potentials/"events" within each ROI. The frequency of these events are typically the units of analysis for downstream biological inference of calcium imaging data, for example, comparing neuronal cell lines from different treatment or genotype groups. We therefore developed a program with an interactive dashboard in MATLAB to produce images that display traces of the average intensity of each ROI across the full video of the green channel. The dashboard allows the user to visualize the fluorescence signal over time and explore these traces within specific timepoints imposed against the frame of the corresponding ROI.

We used this tool to generate "gold standard" events for assessing the accuracy of semi-automated event detection algorithms. We used the traces obtained from the manual cell segmentations with this interactive dashboard to select the beginning of each ROI activity event in each green channel video. For events that were ambiguous or unclear, a subject matter expert was referenced for independent assessment of the events. These event selections

were exported and considered the "gold standard" (or truth) for the frames containing cell activity events for each cell. In this study, the manual event detection was only performed on mean traces that had been smoothed using the DFF method.

2.2.5.2 Motif Correlation

Motif correlation utilizes a set of portions of trace plots from calcium imaging data sets that were manually identified as events. This strategy is similar to methods used to predict human movement using movelets (Bai et al., 2012; He et al., 2014) as well as many other time series tasks (Lin and Shim, 1995; Chan and Fu, 1999; Ge and Smyth, 2000). These portions are referred to as "motifs". The motif correlation method used modified source code from the FluoroSNNAP software package developed by Patel et al. (2015).

This method takes two parameters selected by the user. The first is a correlation threshold, C_t, which is a threshold of similarity measured by the Pearson correlation coefficient between motifs and portions of the trace being analyzed. The second is a height threshold, H_t, which is a threshold of the minimum range of a portion of a trace that can be considered an event. The method also employed a library of motifs that represent activity. For the library of motifs, 23 motifs were generated for this study. Sixteen of these motifs came directly from the FluoroSNNAP software package and 7 additional motifs were generated using traces extracted from other sets of calcium imaging data.

The motifs generated for this method were generated as portions of traces sampled at 10 Hz. Since the calcium imaging data in this set of experiments

was sampled at 4 Hz, we used the MATLAB function `interp1` to linearly interpolate each trace to estimate the trace at 10 Hz (Akima, 1970; Akima, 1974). For each motif, we began with the first timepoint of the interpolated trace being analyzed and extracted the following portion of the trace that spanned the length of the motif. If the range (maximum value minus minimum value) of the extracted portion was larger than H_t, we calculated the Pearson correlation coefficient between the motif and that portion of the trace. If the range of the extracted portion was not larger than H_t, we saved the a zero value in place of a correlation coefficient for that timepoint. This process was repeated for each trace and motif, beginning with every timepoint in the trace. When this method reached a point in the trace where there were fewer timepoints remaining than the length of the motif being considered, the motif was truncated from the end to match the size of the remaining trace. For each trace, we saved the matrix of correlation coefficients with a row for each motif and a column for each timepoint. Then for each trace, we identified the timepoints where the maximum correlation coefficient (across the motifs) was greater than C_t to define timepoints where a cell was active. Then we considered each of these timepoints that was at least ten timepoints (or one second in the interpolated trace) away from the previously identified timepoint as the beginning of an activity event. This step was performed to prevent multiple events being assigned to a single peak. One second was chosen from manual inspection by subject matter experts as a reasonable length of time between peaks in this dataset.

Code for the motif correlation method is found in `findPeaks.m` in the

CIRMIT github repository.

In this study, we applied various values of C_t and H_t to assess performance against manual event detection. For H_t, we considered both static global thresholds and dynamic thresholds that were calculated as a function of each trace's standard deviation.

2.2.5.3 ADEPT

We also considered a second peak-calling algorithm, called adaptive empirical pattern transformation, or ADEPT, which is a statistical method implemented in an accompanying R package Karas et al. (2019). ADEPT was designed to segment walking stride patterns from accelerometry data, but can be implemented to segment any patterns from any time series signal. We therefore hypothesized this may be an effective method to identify activity in fluorescence traces extracted from calcium imaging data. The method also utilizes motifs and computes a similarity statistic chosen by the user across the duration of the series. However, unlike the motif correlation method described above, ADEPT accounts for variability in motif duration and shape. Therefore, it is not necessary to interpolate the calcium imaging trace prior to event detection. In order to account for variability in the motif duration, ADEPT requires a input of a sequence of possible durations of an event and uses this input to scale the motifs, so that for each motif a similarity measurement is calculated for each possible duration of the event.

For each motif, ADEPT calculates a similarity measure for each time point in the trace input. First, ADEPT selects the portion of the trace with the

maximum similarity measure across all of the motifs and motif durations and identifies this portion of the trace as an event. Then ADEPT iteratively selects the portion of the trace with the next largest similarity measure that does not intersect with any of the previously identified events. ADEPT continues this process until the entire trace is segmented. For each segmented portion of the trace, ADEPT saves metrics, such as the index of the motif used to segment that portion of the trace, the duration of the motif, and the value of the similarity measure. In this study, we also saved the range of the portion of the trace segmented and the standard deviation of the full trace to use in downstream analyses.

We used correlation as a similarity measure and assumed all motifs should be between 5 and 30 seconds in length. These values were chosen with input of subject matter experts and from manually inspecting calcium imaging trace plots from this study. We generated input motifs using the same 23 motifs from above, and we also removed up to 40% of the height of each motif from the preceding and following time points. This was done to account for potential events at the beginning or end of a time series experiment, which are hard to measure using correlation-based approaches. We also applied correlation (C_t) and height (H_t) thresholds in the ADEPT method to assess the role the threshold played in the performance of the approach (as compared to manual event detection). However, unlike the previous motif correlation method, these thresholds do not affect the processing steps in ADEPT and were therefore applied after data processing. For H_t, we considered both static global thresholds and dynamic thresholds that were calculated as a function

of each trace's standard deviation.

Code for the ADEPT method, as implemented in this study, is found in `ca_adept.R` in the CIRMIT github repository.

2.2.5.4 Trace Statistics and Threshold Parameters for Each Method

In this study, we applied various combinations of trace statistics (mean vs. quantiles) and threshold parameters (peak amplitude and motif correlation threshold). We applied event detection methods to each of these raw traces, as well as each of these traces smoothed using DFF. We considered motif correlation thresholds from 0.7 to 0.95 incremented by 0.05. In addition, we considered larger motif correlation thresholds from 0.96 to 0.99 by 0.01. For height thresholds, we considered both static global thresholds and dynamic thresholds calculated as a multiple of each individual trace's standard deviation. For the static height thresholds, we considered values from 0.1 to 1 incremented by 0.1. For dynamic thresholds, we considered values from 0.1 to 2 incremented by 0.1. We considered all combinations of these trace statistics and thresholds for each event detection method, and compared the performance of each parameter set to the gold standard of manual segmentation and event detection.

In summary, we employed two event detection methods to identify timepoints of events in each ROI in this study. For each method, we examined several combinations of trace statistics, smoothing procedures, and threshold parameters. In total, we tested 2,400 combinations of trace statistics and parameters for each method, resulting in 4,800 total combinations of event

detection methods and parameterizations. These combinations are illustrated in Figure 2.3.

Figure 2.4 provides an illustration of the entire calcium imaging processing pipeline.

2.3 Statistical Model

For each processing pipeline, we performed regression to jointly estimate the effects pharmacological conditions on the activity rate within each neuron, controlling for cell type or diagnosis (ESC-generated, PTHS or healthy control). We specifically implemented a log-linear mixed effects model with a random intercept corresponding to each experiment.

Let $Y_{i,j}$ be the number of events observed in the j^{th} ROI of the i^{th} experiment. We define t_i as the length of the i^{th} time series video in seconds.

As time may vary across experiments, our outcome of interest is the number of events per time. Thus, we will use $\log(t_{i,j})$ as an offset, with no coefficient being estimated for this variable. For a given experiment $t_{i,j}$, all neurons will likely have the same $t_i = t_{i,j}$. In cases where the cells moves out of the field of view, $t_{i,j}$ will be the number of seconds where any pixel of the cell was in the field of view. Formally, the model we fit was:

$$\log(Y_{i,j}) = b_{0,i} + \log(t_{i,j}) + \beta_0 + \sum_{k=1}^{3} \beta_k I\left(\text{condition}_{i,k}\right) + \sum_{l=1}^{2} \gamma_l I\left(\text{group}_{i,l}\right) + \varepsilon_{i,j} \tag{2.2}$$

where $b_{0,i}$ is a random intercept for i^{th} experiment, $\text{condition}_{i,1} =$ Gabazine,

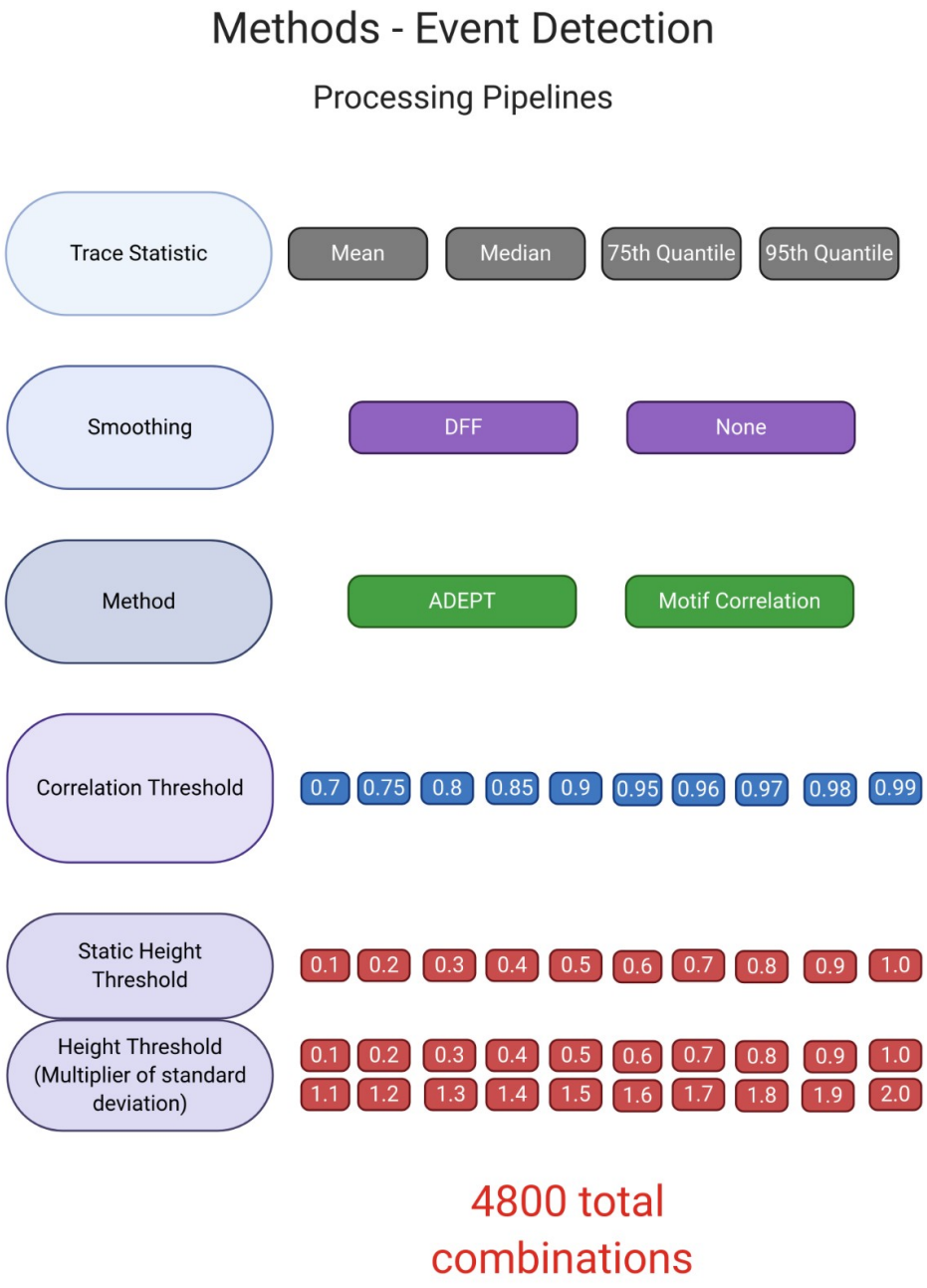


Figure 2.3: Visualization of all event detection methods and parameterizations explored in this study.

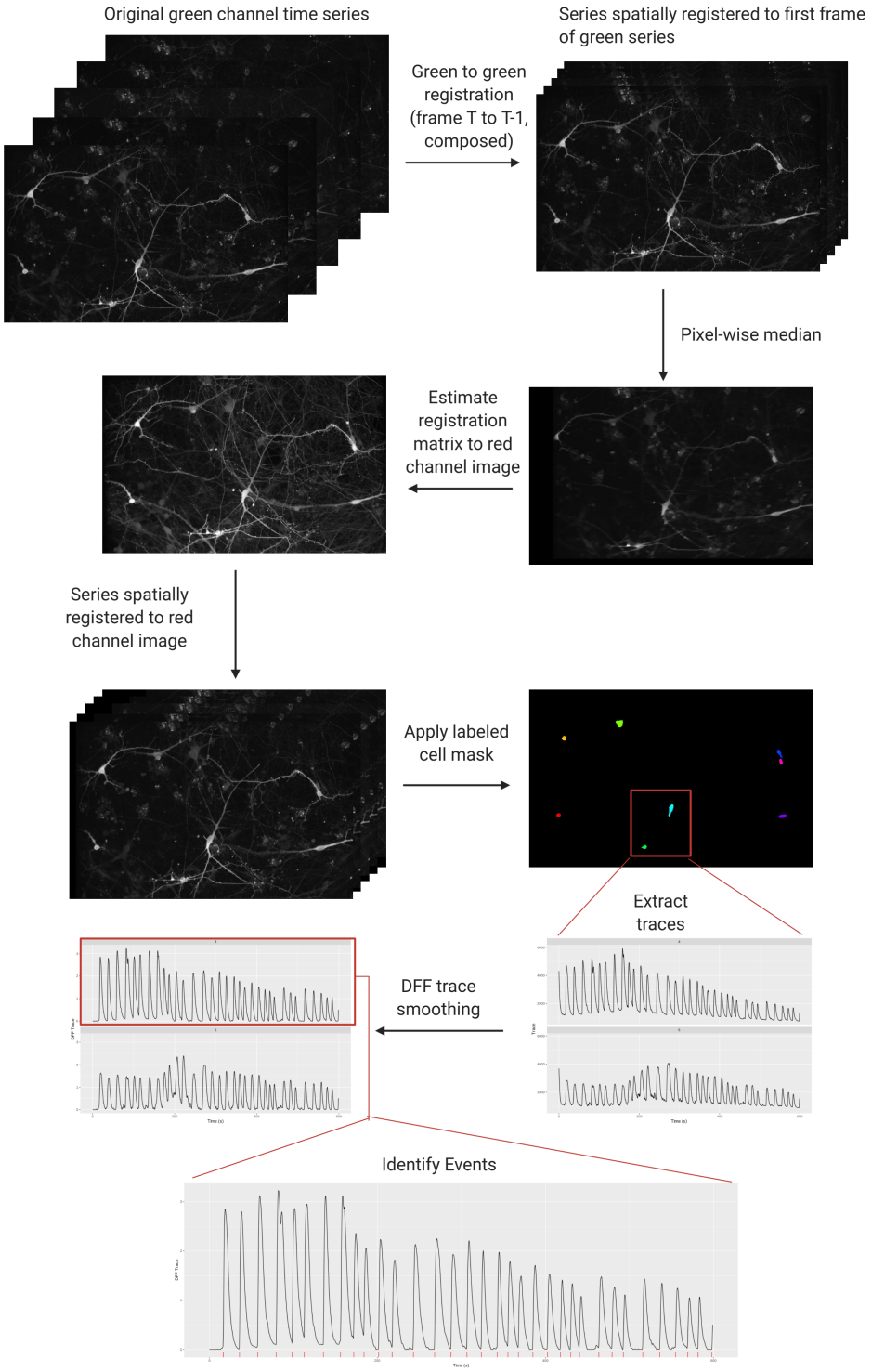


Figure 2.4: Caption on following page.

Figure 2.4: An illustration of the full calcium imaging processing pipeline. We start with the raw calcium imaging time series data. The first step is to register each frame in this series to the space of the first frame. We do this by estimating translation matrices for each frame to the previous frame using a mean square error cost function, then composing these translation matrices to estimate the translation of each frame to the first frame. Next, we take the pixel-wise median of this registered time series and estimate its translation matrix to the red channel image using Matte's mutual information as a cost function. We apply this translation matrix to each frame in the registered series to result in a time series where each frame is registered to the space of the red channel image. After this step, we apply the labeled mask acquired from cell segmentation (Figure 2.1) to extract the trace of a summary statistic for each region of interest in the time series. After this step we smooth the trace of the summary statistic using DFF. Finally, using one of the event detection methods, we identify points in the time series where we estimate the beginning of activity events, denoted in this figure by red tick marks below the trace plot. While in processing, we use the raw intensity images, the $\log_{10}$-transformed images are preferable for visualization.

$\text{condition}_{i,2}$ = DNQXAP5, and $\text{condition}_{i,3}$ = TTX, each compared to the baseline condition $\text{condition}_{i,0}$, and $\text{group}_{i,1}$ = PTHS and $\text{group}_{i,2}$ = ESC, each compared to ROIs from neurotypical subjects. To fit this model we used the R function `glmer` from the `lme4` package (Bates et al., 2015). We specified the model under the `poisson` family for a log link function. Although there may be multiple experiments per donor, we are ignoring this information and treating each experiment as independent, since each experiment is nested within the donors, and we cared about differences in processing strategies at the level of individual ROIs and images. However, we could easily incorporate a random intercept for donor to account for this repeated measure if desired, which may have a larger impact on the effects of diagnosis compared to the effects of drugs. Due to the relatively small sample size from each donor, and our primary interest of estimating the effect of pharmacological intervention, we chose to treat each experiment as an independent observation, rather than

further clustering by donor.

In this study, the coefficients of interest in the model were β_k, for $k = 1, \ldots, 3$. The interpretation of the exponential of each of these coefficients (e^{β_k}) is the relative rate of activity of neurons in this study under condition$_{i,k}$ compared to the baseline condition, controlling for cell group.

In the model, each ROI in a given condition was treated as an observation, with the number of events recorded in a particular condition as the outcome of interest. The model allowed for each group of neurons (controls, PTHS, and ESC) to have their own estimate of average baseline activity rate. The model also allowed for randomness in this baseline activity rate across all unique local fields of view. That is, ROIs from the same recording were estimated to have the same average baseline activity rate, however the model allowed for variation in this estimate between different recordings.

2.3.1 Limiting Models to Active Cells

A portion of the *in vitro* cells in this experiment may have died or otherwise become inactive before calcium imaging was performed. Because of this, we also fit the model above when limiting the population of cells to only those that are determined to be "active", where activity was defined as having at least one event identified by a given method in any condition.

2.4 Processing Evaluation

2.4.1 Evaluating Cell Segmentation Methods

In order to assess the automated cell segmentation methods, we assessed similarities in the resulting ROI/cell mask from each method to the mask created by manual segmentation. We used the Dice similarity coefficient (DSC) (Dice, 1945) as the primary performance metric of interest to assess the mask from the parameter choices for each semi-automated method. The formula for the Dice similarity coefficient is as follows:

$$DSC = \frac{2\,|M_{manual} \cap M_{auto}|}{|M_{manual}| + |M_{auto}|}$$

where M_{manual} is the manually segmented mask, M_{auto} is the automatically segmented mask results from a given set of parameters and $|M|$ is the number of pixels in a given mask. This measure is used as it does not use true negatives (manual and auto both say no cell) in the calulation. A DSC of 1 corresponds to perfect overlap, 0 means no overlap. In this study, there were 17 red channel images. We selected the best automated segmentation method by the highest median Dice similarity coefficient from the data set for downstream performance evaluations.

In addition to the Dice similarity coefficient, we also collected metrics related to false positives, false negatives, and processing time for each automatically segmented procedure. We defined a false positive as an identified cell in the automatic procedure that did not overlap any cells in the manually segmented mask, and a false negative as an identified cell in the manually

segmented mask that did not overlap any cells in the automatic procedure. For each image and method, we considered the proportion of automatically-identified cells/ROIs that were false positives and the proportion of manually identified cells that were false negatives.

All semi-automated methods for cell segmentation used a threshold derived from the standard deviation of the image intensity. In this study, we attempted all methods with a grid of thresholds from 0.5 to 10 image intensity standard deviations, separated by 0.5 standard deviations. All together, we considered the masks resulting from 80 method and parameter combinations in each of 17 red channel images.

2.4.2 Evaluating Event Detection Methods

We evaluated automatic event detection methods against the results of manual event detection for the same corresponding cell and time series. This was accomplished by estimating events in cells identified from the manual cell segmentation using the motif correlation and ADEPT methods described above. For each of the event detection methods, we used a variety of trace statistics and height and correlation threshold parameters. For each method, trace statistic, and set of threshold parameters, we calculated the number of events identified in each cell and time series experiment. Then we compared this result to the number of events in each cell and time series experiment identified by manual event detection. We assessed this comparison using the Pearson and Spearman correlation coefficients, as well as root mean squared error (RMSE).

Additionally, we assessed the frequency of a) inactive cells being falsely identified as active and b) active cells being falsely identified as inactive by each semi-automated method. Therefore, we also calculated and evaluated the following metrics. We defined the false active rate (FAR) as

$$FAR = \frac{N_{FA}}{N_{NA}}$$

where N_{FA} was the number of cells where no events were identified in manual event detection and at least one event was identified in automatic event detection and N_{NA} was the total number of cells where no events were identified in manual event detection.

We defined the false non-active rate (FNR) as

$$FNR = \frac{N_{FN}}{N_A}$$

where N_{FN} was the number of cells where at least one event was identified in manual event detection and no events were identified in automatic event detection and N_A was the total number of cells where at least one event was identified in manual event detection.

We performed all event detection methods on each extracted trace from the manually segmented cells with grids of correlation and height threshold parameters to identify optimal event detection methods and corresponding parameters. We then imposed a tolerance on the FAR and FNR and only considered methods that produce ratios less than or equal to those tolerances. After consultation with subject matter experts, we determined that it was more

important to limit the rate of false active cells in event detection. Therefore, for this study we imposed a tolerance of 0.1 for FAR and 0.25 for FNR. After selecting the subset of trace statistics and parameter thresholds that fall within these tolerances, we selected the top-performing combinations by Pearson correlation coefficient with manually identified events, Spearman correlation coefficient with manually identified events and root mean squared error versus manually identified events.

To select the top performing methods, we treated the calcium imaging data with no pharmacological treatment (i.e. the baseline condition) as a training set and validated the results using the calcium imaging data with pharmacological treatment. We therefore executed the procedure described above on the baseline data, then further assessed method and parameter performance on pharmacologically-treated data. These conditions were thought to either increase or decrease activity.

We will also assess event detection methods by comparing the results of the statistical models described above across processing pipelines. We will consider the manual event detection method to be the best estimate of the truth, and compare the results of the statistical models run on all other processing pipelines to the results from the manual event detection. We assessed each processing pipeline using similarity of summary statistics of metrics of interest to those produced by manual event detection, as well as how similar results and inference from the statistical models were to those produced by manual event detection. This allowed us to assess the biologically-relevant statistical inference that we drew from an automated pipeline and compare it to the

inference that we drew from the manual pipeline.

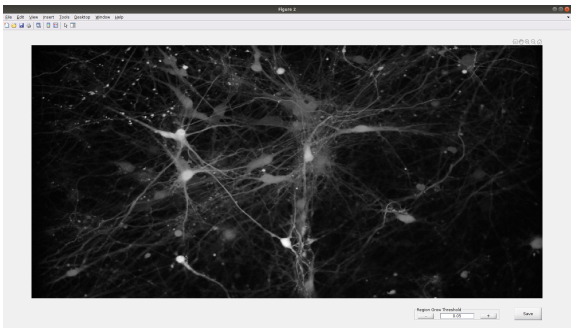
2.5 CIRMIT

We created an interactive data visualization tool using MATLAB to improve visualizing calcium imaging data and performing manual event detection. The tool is named the Calcium Imaging Real-time MATLAB Interactive Toolbox, or CIRMIT. This tool is a GUI-based software product that allows data processing steps described above to implemented on one red channel image and corresponding green channel time series at a time. CIRMIT has been compiled using MATLAB Runtime and is available as a standalone application. The user select pairs of red channel and green channel CZI image files, and ROI and fluorescence data is automatically processed and saved. The software then displays an interactive dashboard allowing the user to visualize the processed calcium imaging data. The dashboard can "play" the time series and closely visualize up to four cells at a time. The dashboard also displays the ROI mask, the full time series image, a trace plot, and zoomed-in footage of the selected cells. As the dashboard plays, the full time series, zoomed-in footage, and a uniquely colored marker for each cell on the trace plot updates with each frame. This allows the user to monitor the time series and assess the quality of automatically identified events. The user can also add and remove events for each trace and save the event data. In addition, this interactive dashboard can be launched with previously processed and saved data.

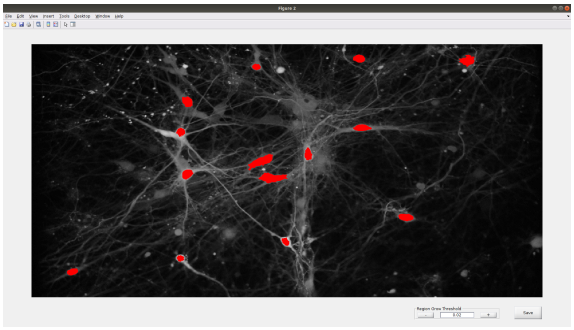
Source code for CIRMIT is found in the CIRMIT github repository. An

installer for a standalone application compiled for windows (.exe) is also available in the repository in the `./compiled/Windows/CIRMIT/for_redistribution` folder.

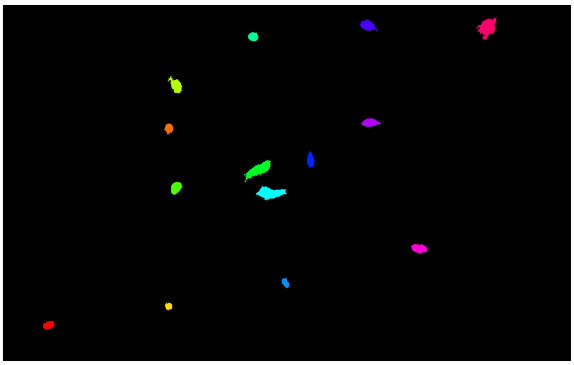
CIRMIT was the interface used to create the manual cell segmentation and manual event detection data as described above. We used the option for manual cell segmentation that displayed the $\log_{10}$ transformed red channel image and allowed the user to point and click to identify events as describe in section 2.2.1.1 above. An image of the dashbard for manual cell segmentation in CIRMIT is displayed in Figure 2.5. The manual event detection was performed as described in section 2.5.1.1 above using the interactive dashboard. An image of the dashboard for manual detection in CIRMIT is displayed in Figure 2.6.



(a) Window displaying $\log_{10}$-transformed red channel image



(b) User identifies regions of interest by point-and-click



(c) Labeled mask saved

Figure 2.5: Visualization of manual cell segmentation using CIRMIT. In panel (a), we see the user is presented with the $\log_{10}$-transformed red channel image. In panel (b), we see the ROIs superimposed on the image in red. These regions of interest are displayed when the user identifies a cell by pointing-and-clicking. A threshold value can be toggled in the bottom right and previous regions can be deleted by pointing-and-clicking. A "Save" button in the bottom right corner allows for the user to save a labeled mask. In panel (c), we see the result of saving the binary mask, with each unique ROI identified by a unique color.

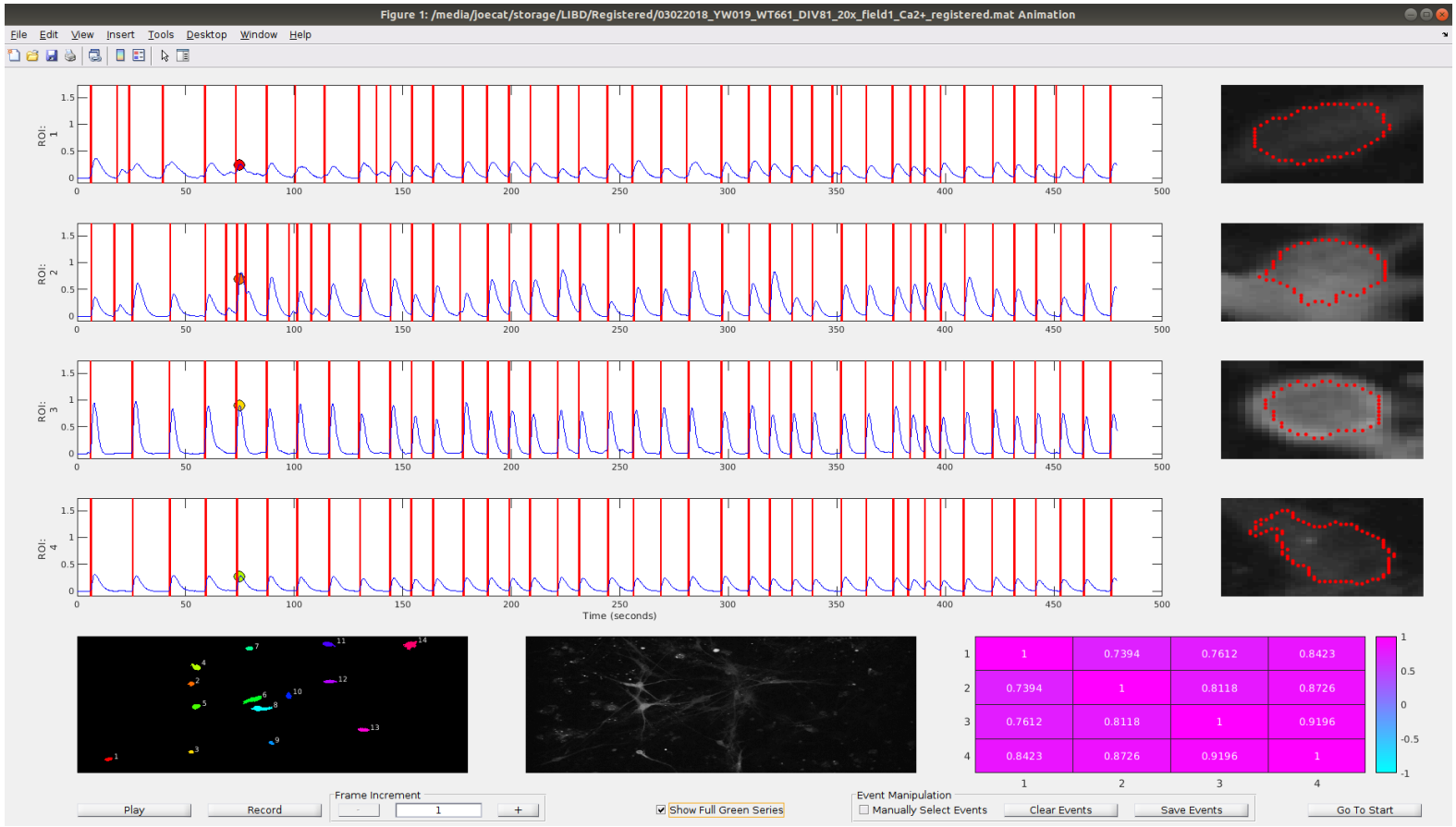


Figure 2.6: A screenshot of the main figure produced by CIRMIT. We see DFF-smoothed trace plots from four regions of interest with a zoomed in image of each region of interest to the right of the trace plot. The round marker on each trace plot indicates what time in the series the figure is displaying. The red lines on the trace plots indicate where events were identified. Events can be added and removed, then saved by the user by using functionality at the bottom of the figure. The bottom left image displays the labeled mask used in this experiment. The bottom middle image displays the full calcium series recording. This image also be toggled to display the static red channel image. The bottom right displays a correlation matrix with heatmap for the four traces being displayed. Using functionality at the bottom of the figure, the user can "Play" the figure with images and the correlation matrix updating frame-by-frame. The user can also write this figure to an AVI video format to be saved and played later.

Chapter 3

Results

We evaluated the effects of automated processing methods and parameter choices on calcium imaging data. This study was conducted on calcium imaging data collected from human iPSC- and ESC-derived neurons across seven cell lines in multiple pharmacological conditions (Table 2.1). In total, we considered 80 cell/ROI segmentation and 4,800 peak detection methods and parameter choices, and compared the fidelity of each approach against a "gold standard" dataset of manual segmentation and peak detection by trained neuroscientists.

3.1 Cell segmentation performance was variable across methods and parameters

We evaluated each cell segmentation method and parameter choice against the manual cell segmentation data using the *DSC*. We specifically compared the automatically-segmented binary region of interest (ROI) mask to the corresponding binary ROI mask generated from manual cell segmentation at

the resolution of individual pixels for four different methods across a range of standard deviation thresholds (Figure 3.1). We found a wide range of *DSC* values across all four methods and 20 thresholding parameters.

In general, the two median-based methods ("Median-Circles" and "Median-Eccentricity") were more highly dependent on the parameter choice than the region growing-based methods, with low *DSC* values for low thresholds. The region-growing methods showed relatively consistent performance across the range of thresholding values, with region growing-circles performing best as low thresholds and region growing-eccentricity at high thresholds. The two eccentricity-based methods tended to show better performance at high thresholds (> 4) than the circle-based methods. Misspecifying the parameter choice (without the available manual segmentation) therefore has the largest impact on the median based methods. Lastly, we note that none of the methods showed median $DSC > 0.5$ in this study, which corresponds to only moderate accuracy.

We further considered the error rates across the top-performing (by median *DSC*) threshold for each of the four methods (Table 3.1). While the median-eccentricity method with threshold 7 had the largest median *DSC* of all the methods, optimized parameters for each methods showed largely comparable accuracies (although the optimal threshold for each method was variable).

We further assessed the distribution of *DSC* across images to assess the variation of accuracies using the median best-performing threshold for each method (Figure 3.2). Here, while each method produced similar median *DSC* value, the variance across the entire dataset was less in the median-eccentricity

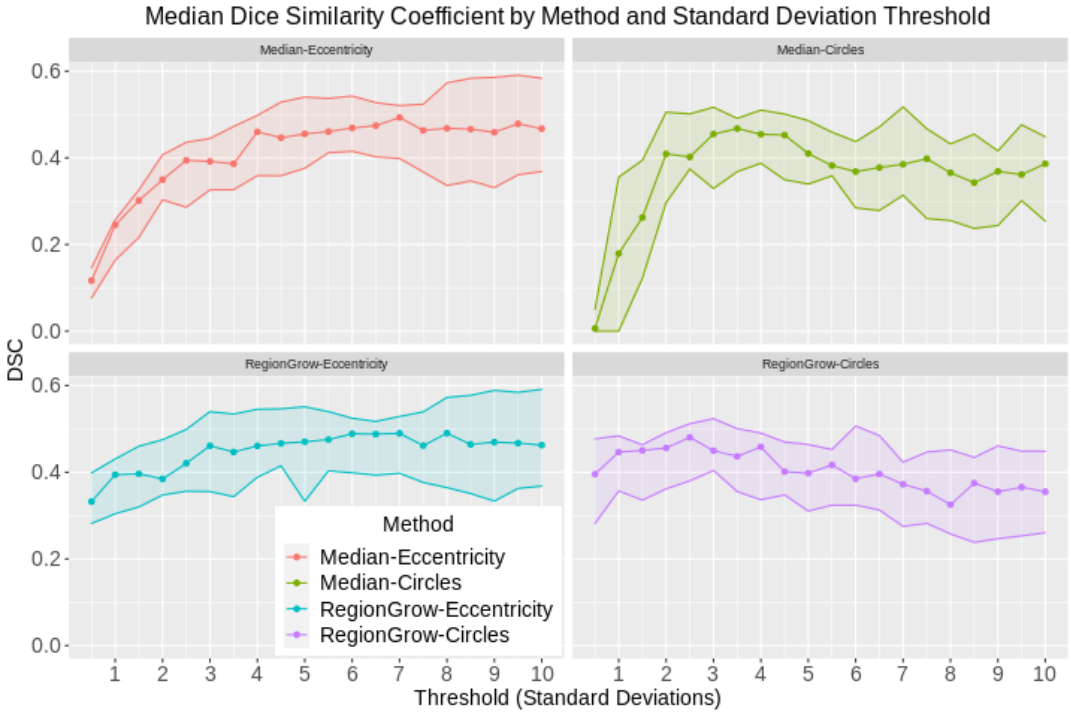


Figure 3.1: Median *DSC* vs. Threshold for each Method Compared to the Manual Segmentation. Here we present different thresholds of the red channel image, based on a multiplier of the standard deviation of that image (e.g. "7" means $7 \times \sigma$) and the Dice similarity coefficient (*DSC*) comparing the manual segmentation to the segmentation based on each method. The points in this figure display the median *DSC* compared to manual cell segmentation for each method at each threshold. The ribbons surrounding the points display the interquartile range (IQR) of each method at each threshold. We see that all methods perform relatively equivalently around a threshold of 4σ. Improvements can be made with higher thresholds for those methods without circle-finding.

Method	Threshold	Median *DSC*
Median-Eccentricity	7.0	0.493
Region Growing-Eccentricity	8.0	0.490
Region Growing-Circles	2.5	0.480
Median-Circles	3.5	0.468

Table 3.1: Top-performing standard deviation-based threshold for each cell segmentation method. In this figure we assess the top-performing threshold as that with the minimum median *DSC* compared to manual cell segmentation.

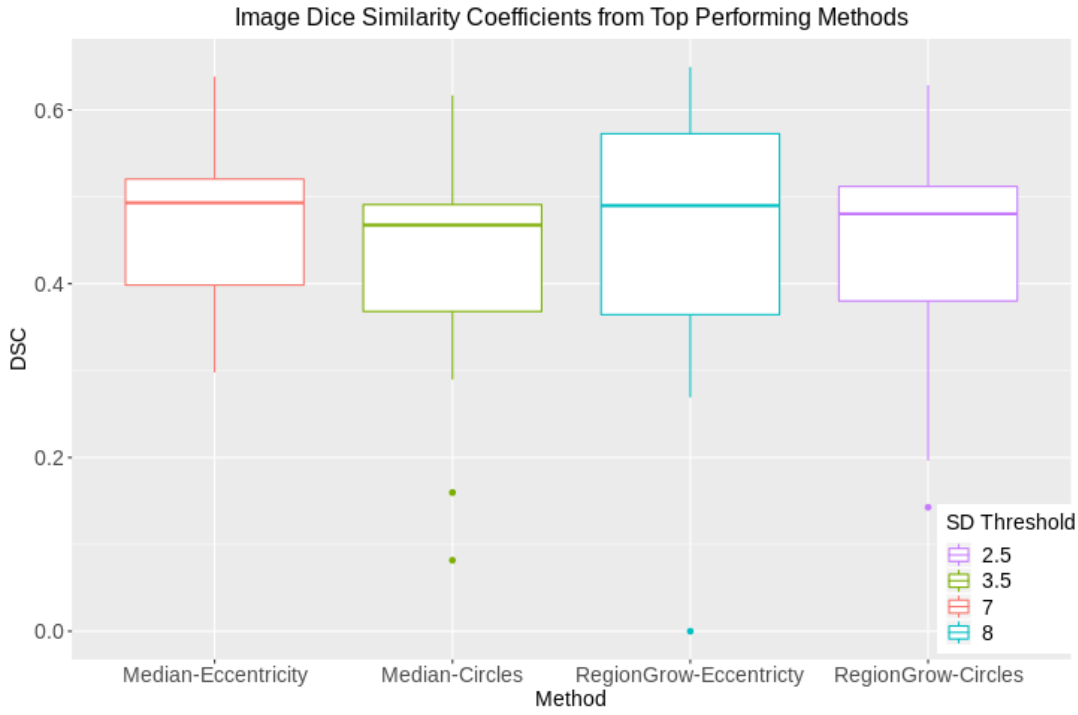


Figure 3.2: Boxplots of image *DSC* vs. manually segmented mask for each method's top-performing threshold. x-axis: the four cell segmentation methods, colored by their standard deviation-based threshold. y-axis: boxplot with *DSC* scores for each of the red channel images in the dataset for that method compared to manual cell segmentation. In this figure, we specifically examined the top-performing threshold by median *DSC* score for each method.

method with threshold 7. This method also provided the largest minimum *DSC* in the dataset of these selected methods, suggesting the approach was fairly consistent across all images.

We were also interested in how well each method correctly identified ROIs as cells. In order to evaluate this, we calculated ratios of false positives and false negatives for each automated method and parameter choice. In Table 3.2, we defined *FP* as the ratio of automatically identified cells that did not overlap with any manually identified cells and defined *FN* as the ratio of manually identified cells that did not overlap with any automatically identified cells.

Method	Threshold	Median FP	Median FN
Median-Eccentricity	7.0	0.286	0.455
Region Growing-Eccentricity	8.0	0.250	0.500
Region Growing-Circles	2.5	0.143	0.429
Median-Circles	3.5	0.231	0.455

Table 3.2: Median false positive (FP) and false negative (FN) ratios for each method's top-performing threshold. In this study, FP is the ratio of automatically identified cells that do not intersect with a manually identified cell and FN is the ratio of manually identified cells that do not intersect with an automatically identified cell.

All methods generated a median $FP < 0.3$ and a median $FN \leq 0.5$. It was not surprising that the automated cell segmentation resulted in a larger ratio of false negative cells, as manual cell segmentation allowed for a expert user to identify cells that did not fluoresce very brightly, and thus did not provide adequate signal to be identified in an automated method. However, moderately high ratios of false positive cells was somewhat concerning, suggesting that improvements in the cell refinement stage of automated cell segmentation may be of interest for future work.

We next examined computational considerations in implementing these methods, particularly to large datasets of many recordings. Table 3.3 displays the median processing time for each of the top-performing methods, which shows that the methods using the median thresholding were much faster to execute - by two orders of magnitude - than the region growing methods.

We were interested in further examining the difference in distributions of DSC scores between the four methods when the same standard deviation-based threshold was applied. As seen in Figure 3.1, all four cell segmentation methods produced very similar median DSC scores with threshold 4. We selected this threshold in order to further examine the distribution of DSC

Method	Threshold	Median Proc Time (s)
Median-Eccentricity	7.0	0.388
Region Growing-Eccentricity	8.0	334.365
Region Growing-Circles	2.5	358.808
Median-Circles	3.5	0.596

Table 3.3: Median processing times for each top-performing cell segmentation method. We see that the median thresholding methods perform much faster than the region growing based methods.

scores compared to manual cell segmentations with a comparable threshold value where all methods performed moderately well. Figure 3.3 displays the boxplots for *DSC* for all 17 red channel images in the dataset for each method with threshold 4. While each of the methods produced a similar median *DSC*, the variance of the *DSC* scores across the entire dataset differed by method. Furthermore, while the median *DSC* scores were extremely similar across the methods at this threshold, the distribution and outliers of *DSC* scores can differ across the methods.

Figure 3.1 shows that the region growing and median-eccentricity methods produced very similar median *DSC* scores at the larger standard deviation-based thresholds. The top-performing overall method by this metric was the median-eccentricity method with threshold 7. While the top-performing region growing-eccentricity method was that with threshold 8, the region growing-eccentricity method with the second largest median *DSC* (and third largest overall) was that with threshold 7 (median $DSC \approx 0.490$). We are interested in assessing how similar the binary masks produced with the top median-eccentricity method are to the region growing-eccentricity method with the same threshold. Figure 3.4 shows the boxplot of each of the 17 red

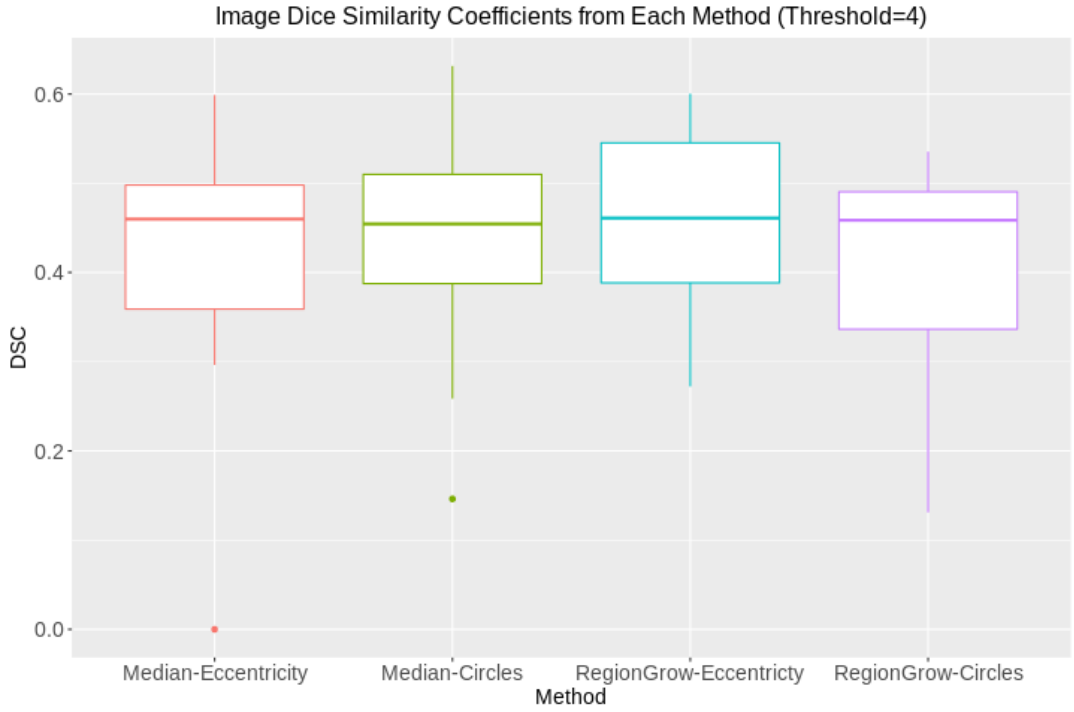


Figure 3.3: Boxplots of image *DSC* vs. manually segmented mask for each method with threshold equal to 4. x-axis: each of the four cell segmentation methods, colored by their standard deviation-based threshold. y-axis: *DSC* scores for each of the red channel images in the dataset for that method compared to manual cell segmentation. In this figure, we examine the same threshold for each method. We chose to examine a threshold equal to 4, because at this threshold, all methods perform similarly by median *DSC* score, as illustrated in figure 3.1

channel images' *DSC* comparing the median-eccentricity mask against the region growing-eccentricity mask with threshold 7. The masks produced with these two methods at threshold 7 were extremely similar for all 17 red channel images, with a minimum *DSC* greater than 0.93 (compared to each other, not the manual segmentation). This supports that, at near-optimal threshold values, the median-eccentricity and region growing-eccentricity methods produced extremely similar cell segmentations in this dataset.

The median-eccentricity method and region growing-eccentricity method were both relatively stable at large standard deviation-based thresholds, and tended to produce very similar cell segmentations with like threshold values. Since the top-performing median-eccentricity method produced cell segmentations with less variance in *DSC*, and was much faster in processing time than the region growing-eccentricity method, we selected the median-eccentricity method with threshold 7 for the automated pipelines in this experiment. Figure 3.5 displays a visualization of the median-eccentricity mask with threshold 7 compared to the manually segmented mask. In the panel (a), the $\log_{10}$ transformed red channel image with the median *DSC* score from this method is displayed. In panel (b), a visualization of the automatically segmented mask overlapping the manually segmented mask. In the panel (c), the pixel-wise 99^{th} quantile of the registered calcium imaging time series for this experiment is displayed, which demonstrates which cells provided the most detectable activity (the brightest intensities). The automatically segmented mask identified 3 of the 7 cells identified by manual segmentation and also identified 2 other ROIs as cells that were not selected in manual segmentation. The bottom

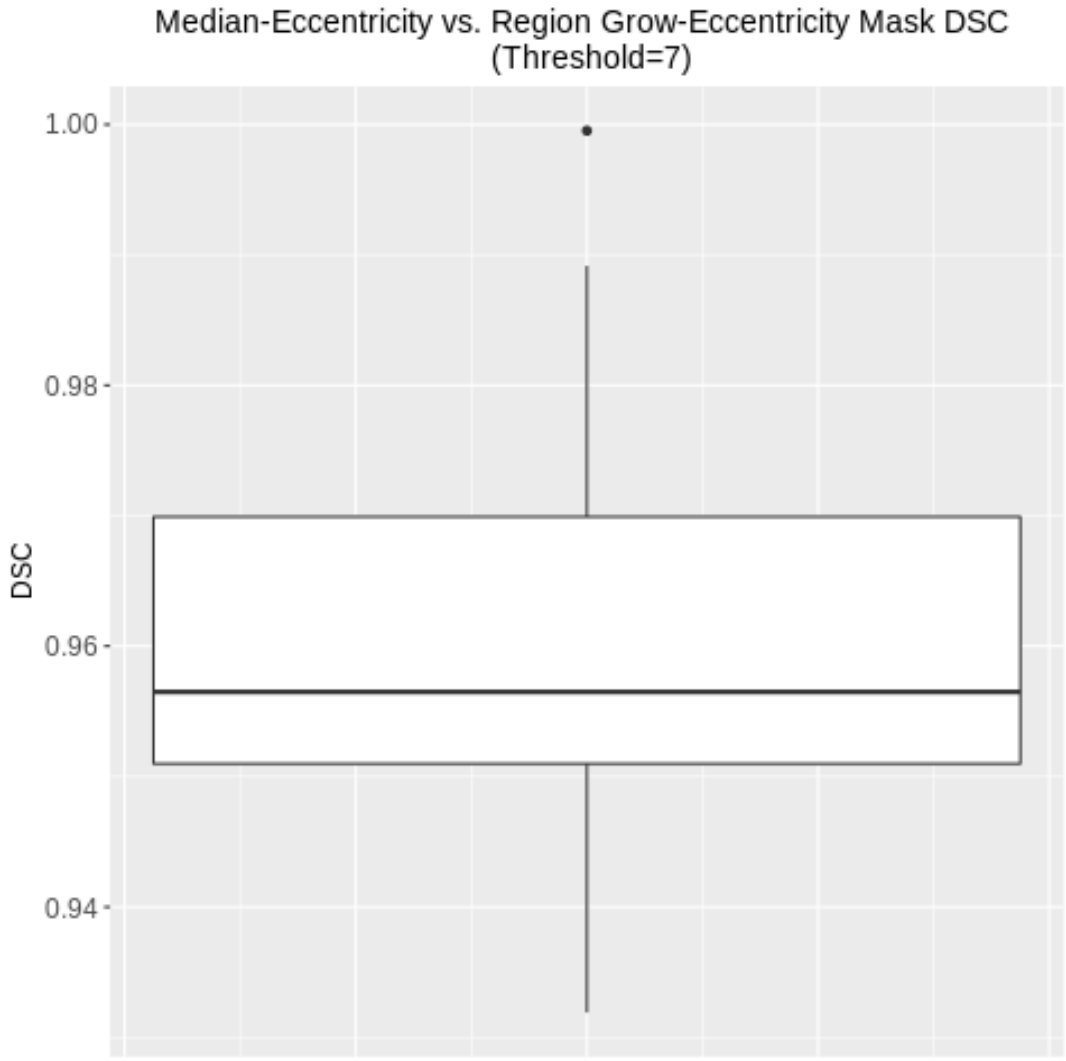
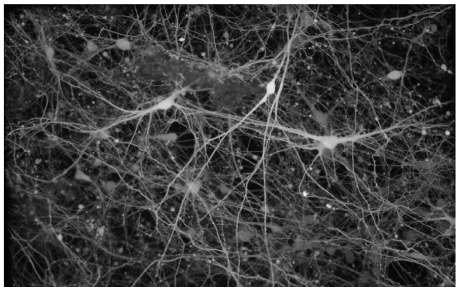
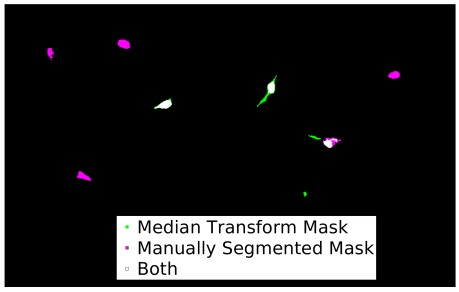


Figure 3.4: Boxplot of *DSC* scores for median-eccentricity vs. region growing-eccentricity with threshold 7. We saw in figure 3.1 that at larger threshold values, the median-eccentricity and region growing-eccentricity methods performed very similarly by median *DSC*. While, region growing-eccentricity with a threshold of 8 resulted in the largest median *DSC* for the region growing-eccentricity method, and second largest median *DSC* overall, the region growing-eccentricity method with threshold 7 resulted in the third largest median *DSC* overall. In this figure, we examine the the similarity between cell segmentations generated by the median-eccentricity and region growing-eccentricity methods with a like threshold of 7. We see that across all red channel images in the set, both methods create very similar masks.

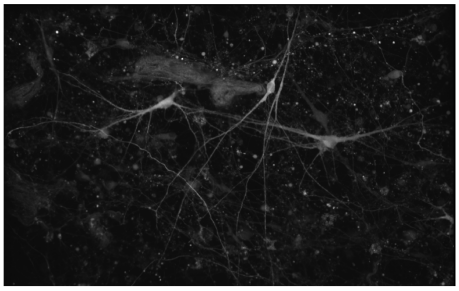
figure shows that the 3 cells that both methods selected did appear to have the brightest intensities in the calcium imaging time series.



(a) Red channel image



(b) Dice visualization



(c) Pixel-wise 99^{th} quantile of calcium imaging time series

Figure 3.5: Visualization of *DSC* for the median case of the overall top-performing method. The overall top-performing method was the median-eccentricity method with standard deviation-based threshold 7. In panel (a), we see the red channel image. In panel (b), we see the manually segmented mask and automatically segmented mask superimposed on each other. Black pixels indicate where neither method identified a cell. Magenta pixels indicate where only the manual segmentation identified a cell. Green pixels indicated where only the median-eccentricity automatically segmented mask identified a cell. White pixels indicate where both methods identified a cell. In panel (c), we see the pixel-wise 99^{th} quantile of the registered calcium imaging time series. That is, each pixel in the image displays the 99^{th} quantile intensity for that pixel across the entire length of the time series. Panels (a) and (c) display the $\log_{10}$ transform the described. This is for visualization purposes.

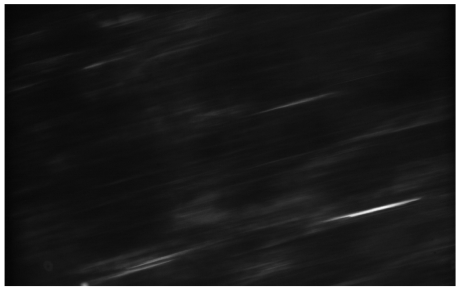
3.2 Registration methods can remove drift in calcium imaging data

Figure 3.6 displays a visualization of of the registration method used in this study. All images in this figure show the $\log_{10}$ transformed intensities. In panel (a), the pixel-wise median of the unregistered time series across the time axis is displayed. This image shows a blurry image as there is significant drift in this time series spatially. In panel (b) the pixel-wise median of the time series registered to the first frame of the time series is displayed, which shows that this novel registration can result in an aligned image. In the panel (c), the red channel image is displayed. In panel (d), the pixel-wise median of the time series registered to the red channel image is displayed. This figure displays that even in cases with significant drift, the image registration technique employed in this study produced stable estimates of image translation, and was successful in approximately registering frames from a time series to the space of the red channel image.

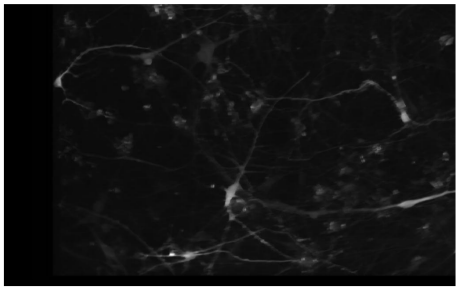
3.3 Event detection: optimized parameterizations produced results comparable to manual event detection

3.3.1 DFF and static height threshold produced viable methods for optimized parameterizations

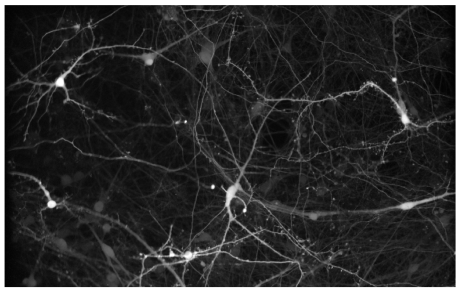
To select viable peak detection methods, we applied each method, trace statistic and threshold parameters to each manually segmented cell in the dataset.



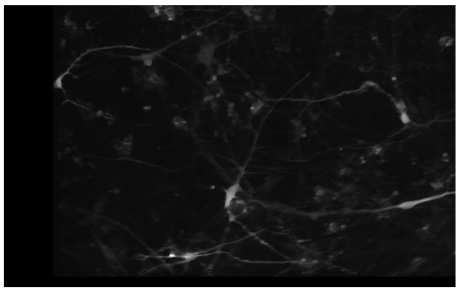
(a) Raw series pixel-wise median



(b) Series registered to frame 1 pixel-wise median



(c) Red channel



(d) Pixel-wise median of series registered to red channel

Figure 3.6: Visualization of time series registration. In panel (a), we see the pixel-wise median of the unregistered calcium imaging time series. At each pixel, we see the median intensity across the entire length of the time series. In panel (b), we see the pixel-wise median of the calcium imaging time series registered to the first frame of the series. In panel (c), we see the red channel image. In panel (d), we see the pixel-wise median of the calcium imaging time series registered to the space of the red channel image. All images in this figure display the log base 10 transformed versions of the described images. While in processing, we use the raw intensity images, the $\log_{10}$ transformed images are preferable for visualization.

Then, for each time series, we calculated the total number of events identified by the method in each ROI and compared this to the number of events manually identified in that ROI. Using only the ROIs in their baseline condition (no pharmacological treatment), we retained all trace statistic and threshold parameter combinations with a $FAR \leq 0.1$ and a $FNR \leq 0.25$. In this dataset, at baseline conditions, there were 119 unique ROIs, of which 34 had no events (and 85 had at least one event) identified in manual peak detection. Therefore, we only consider automatic peak detection methods where less than or equal to 3 ROIs were falsely identified as having at least one event and less than or equal to 21 ROIs were falsely identified as having no events.

In this study, we applied 2400 combinations of trace statistics and threshold parameters for each method. Under the motif correlation method, 63 (2.625%) of these combinations produced results below the pre-specified FAR and FNR tolerances. Under the ADEPT method, 80 (3.333%) of these combinations produced results below the pre-specified FAR and FNR tolerances. The relatively few number of parameterizations that fell within these tolerances suggests that selecting proper threshold parameters was very important to avoid significant in falsely identifying ROIs as active or inactive. Table 3.4 displays how many of these viable parameterizations used DFF traces.

All viable parameterizations in the ADEPT model used a DFF trace, while 54 of the 63 viable motif correlation parameterizations used a DFF trace (Table 3.4). Here, all viable parameterizations for both methods that used a DFF trace also used a global static height threshold. The 9 motif correlation parameterizations that did not use a DFF trace all used a dynamic height threshold

Method	Viable Parameterizations (N=2400)	Used DFF (%)
Motif Correlation	63	54 (85.7%)
ADEPT	80	80 (100%)

Table 3.4: Table of viable parameterizations that used DFF traces. For each method, we display the total number of viable parameterizations. A viable parameterization was considered to be one that resulted in a $FAR \leq 0.1$ and $FNR \leq 0.25$. In addition, we display the number and percentage of viable parameterizations that used a DFF smoothed trace. It should be noted, that in all of the viable parameterizations that used a DFF trace, a global static height threshold was also used. We see that every viable ADEPT parameterization and all but 9 viable motif correlation parameterizations used a DFF trace.

based off the trace standard deviation. From this result, we recommend using a DFF-smoothed trace with a static height threshold for processing future calcium imaging datasets using our automated pipeline.

In this study, we examined the Pearson (ρ_P) and Spearman (ρ_S) correlation coefficients, as well as root mean squared error (RMSE) to assess each method's agreement with manual peak detection. For each of these metrics, we chose the top performing trace statistic and threshold parameterizations for each method (and all of the top performing combinations used a DFF trace and static height threshold as above). In this results section, we refer to the correlation threshold as C_t and the height threshold as H_t.

There was generally agreement between all three metrics for the motif correlation method (Table 3.5). The top-performing parameterization involved the DFF-smoothed mean trace with correlation threshold 0.9 and global static height threshold 0.2. The top-performing parameterizations for the ADEPT method involved the DFF-smoothed 75^{th} quantile trace with global static height threshold 0.2, with correlation thresholds that varied by metric. Pearson correlation coefficient and root mean square error agreed with a correlation

Method	Trace	C_t	H_t	ρ_P	ρ_S	RMSE	FAR	FNR
Best performing by Pearson correlation coefficient								
Motif Correlation	Mean	0.90	0.2	0.725	0.816	11.9	0.088	0.047
ADEPT	Q75	0.85	0.2	0.763	0.818	10.9	0.088	0.071
Best performing by Spearman correlation coefficient								
Motif Correlation	Mean	0.90	0.2	0.725	0.816	11.9	0.088	0.047
ADEPT	Q75	0.98	0.2	0.760	0.821	11.4	0.088	0.071
Best performing by root mean squared error								
Motif Correlation	Mean	0.90	0.2	0.725	0.816	11.9	0.088	0.047
ADEPT	Q75	0.85	0.2	0.763	0.818	10.9	0.088	0.071

Table 3.5: Table of top-performing methods and parameterizations by Pearson correlation (ρ_P), Spearman correlation(ρ_S), and root mean squared error (RMSE). We display the trace statistic, correlation threshold (C_t) and height threshold (H_t) for the top-performing parameterization by each metric. We see that, for the motif correlation method all three metrics are optimized by the same parameterization. For the ADEPT method, we see all top-performing parameterizations utilizing the 75^{th} quantile trace and $H_t = 0.2$. For the ADEPT method, ρ_P and RMSE are optimized by the same parameterization with $C_t = 0.85$, while ρ_S is optimized by a parameterization with $C_t = 0.98$.

threshold of 0.85, while Spearman correlation coefficient favored a higher correlation thresholds with 0.98.

Among the viable parameterizations, the most common global static height thresholds for both methods were 0.2 and 0.3. Figure 3.7 displays the Pearson correlation coefficient between automatically identified events and manually identified events plotted versus motif correlation threshold, faceted by height threshold and method. The Pearson correlation coefficient of automatically identified events compared to manually identified was less sensitive to the correlation threshold in the ADEPT method than it was in the motif correlation method. In the ADEPT method, identifying events from the 75^{th} quantile trace with $H_t = 0.2$ consistently outperformed other trace statistic and height threshold combinations. The motif correlation method showed that, although

the mean trace with $H_t = 0.2$ and $C_t = 0.9$ had the largest Pearson correlation coefficient, the median trace with the same thresholds performed very similarly. Additionally, the 75^{th} quantile trace with $H_t = 0.3$ and $C_t = 0.7$ also produces a Pearson correlation coefficient above 0.7.

After inspection of the results of the viable parameterizations of each method, we decided to select three parameterizations of each method for further analysis. In addition to the parameterizations in Table 3.5, we selected additional parameterizations based on inspection of Figure 3.7. Each of the methods used a DFF-smoothed trace and global static height threshold. The parameterizations, with relevant performance metrics, are displayed in Table 3.6. For future referencing, we assign each model and parameterization with a identification label letter (A through F), shown in the "ID" column of Table 3.6.

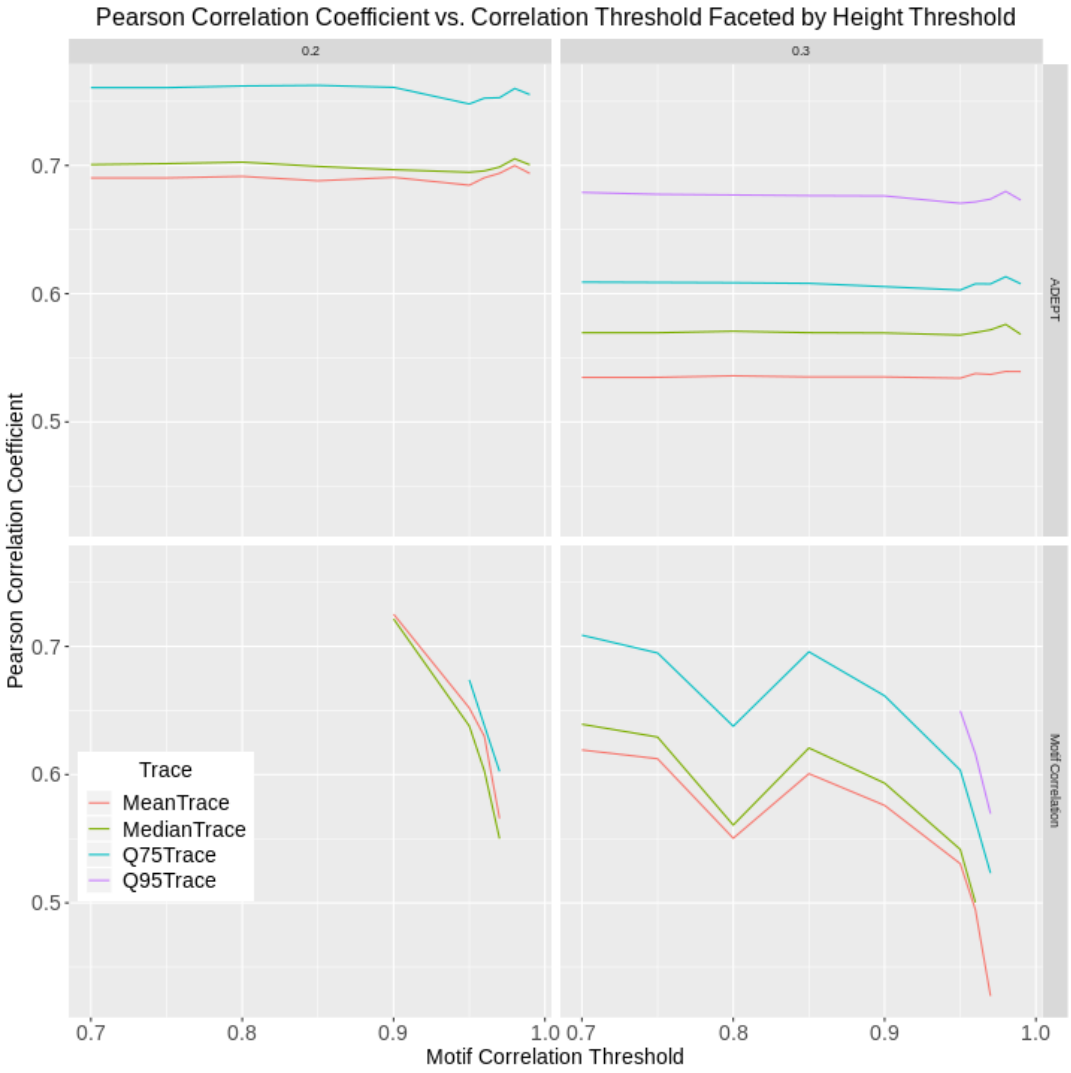


Figure 3.7: Caption on following page.

Figure 3.7: Pearson correlation vs. C_t faceted by H_t and method. Upon investigating the viable parameterizations, we discovered the most common global static height thresholds (H_t) for both methods were 0.2 and 0.3. For each of these values of H_t, we display the Pearson correlation coefficient between viable parameterizations for each automated peak detection method and manual peak detection versus the correlation threshold (C_t). The left column displays these plots for $H_t = 0.2$ and the right column displays these plots for $H_t = 0.3$. The top row displays the Pearson correlation coefficients for the viable ADEPT parameterizations and the bottom row displays the Pearson correlation coefficients for the viable motif correlation parameterizations. In the figure, the trace statistic used in the parameterization is represented by color. We see that the ADEPT method is less sensitive to different values of C_t. That is, within each combination of trace statistic and H_t, we observe a similar Pearson correlation coefficient across all values of C_t. We also observe that for the ADEPT method, the 75^{th} quantile trace with $H_t = 0.2$ consistently outperforms other parameterizations. In the motif correlation method, we see that the value of C_t does have a more noticeable affect on the Pearson correlation coefficient. As this figure only displays viable parameterizations, the data displayed is only from parameterizations that resulted in FAR and FNR under the thresholds stated earlier. That is why some combinations of trace statistic, H_t, and C_t are not displayed.

We also evaluated the performance of these methods and parameterizations for calcium imaging data from ROIs in pharmacologically treated conditions (Table 3.7). As we are ultimately interested in applying a generalized linear model to these data, we were particularly interested in ρ_P as a metric to evaluate performance. Note that for the pharmacologically manipulated set of data, ID B provides the largest ρ_P among the motif correlation parameterizations and ID E provides the largest ρ_P among the ADEPT parameterizations.

Figure 3.8 displays a scatter plot the number of manual events identified versus the number of events automatically identified faceted by the 6 methods and parameterizations described in Tables 3.6 and 3.7 for ROIs in the baseline condition. All methods and parameterizations tended to underestimate events

ID	Method	Trace	C_t	H_t	ρ_P	ρ_S	RMSE	FAR	FNR
A	Motif Correlation	Mean	0.90	0.2	0.725	0.816	11.9	**0.088**	**0.047**
B	Motif Correlation	Median	0.90	0.2	0.721	0.805	12.1	**0.088**	**0.047**
C	Motif Correlation	Q75	0.70	0.3	0.709	0.793	12.3	**0.088**	0.082
D	ADEPT	Q75	0.85	0.2	**0.763**	0.818	**10.9**	**0.088**	0.071
E	ADEPT	Q75	0.98	0.2	0.760	**0.821**	11.3	**0.088**	0.071
F	ADEPT	Q75	0.70	0.2	0.761	0.818	11.0	**0.088**	0.071

Table 3.6: Table of top performing methods to examine. After examination of table 3.5 and Figure 3.1, we decided to further examine 3 parameterizations from each peak detection method, displayed in this table. The bold values in the table represent the optimal value of the metric displayed in the column among these 6 models. We also include an "ID" column where we give each method and parameterization a unique letter ID (A through F) that will be used for future reference.

ID	Method	Trace	C_t	H_t	ρ_P	ρ_S	RMSE	FAR	FNR
A	Motif Correlation	Mean	0.90	0.2	0.613	0.906	6.98	**0.007**	**0.105**
B	Motif Correlation	Median	0.90	0.2	**0.617**	0.895	**6.96**	**0.007**	0.116
C	Motif Correlation	Q75	0.70	0.3	0.589	0.866	7.06	**0.007**	0.163
D	ADEPT	Q75	0.85	0.2	0.575	0.908	7.13	**0.007**	0.116
E	ADEPT	Q75	0.98	0.2	0.595	0.907	7.05	**0.007**	0.116
F	ADEPT	Q75	0.70	0.2	0.577	**0.912**	7.12	**0.007**	**0.105**

Table 3.7: Table of top performing methods to examine results on calcium imaging experiments with pharmacological treatments. As recommended by subject matter experts, we selected the 6 methods from calcium imaging data in the baseline condition. However, we also wished to assess the performance of each method on the data calcium imaging data where activity is manipulated by pharmacological intervention. The bold values in the table represent the optimal value of the metric displayed in the column among these 6 models. We see that in general, the methods consistently produced larger Spearman correlation coefficients (ρ_S), lower root mean squared error (RMSE), and lower false active rate (FAR) on this data with pharmacological intervention when compared to manual event detection than on data in the baseline condition. However, the methods also consistently produced lower Pearson correlation coefficients (ρ_P) and larger false non-active rates (FNR) than on data in the baseline condition.

in ROIs where a large number of events were manually identified. We also note that most of these cases occur in ROIs from healthy control donors. This is potentially problematic, as it could result in bias in the estimate of relative activity rate for ROIs from the PTHS and ESC groups compared to controls. It is preferable to select methods and parameterizations that consistently over- or underestimate events across groups. Visually inspecting this figure, it appears that IDs A, B, and E tended to underestimate the number of events for a given ROI regardless of group. Figure 3.9 displays the same results from event identification in the pharmacologically-induced conditions. We observed that in this data, the automatic event identification methods also tended to underestimate events in highly active ROIs. Furthermore, IDs A, B, and E also tended to underestimate the number of events for a given ROI regardless of group in the pharmacologically-induced conditions.

We selected one parameterization from each automated peak detection method for more in-depth analysis. For the motif correlation method, we selected ID B (DFF-smoothed median trace with $H_t = 0.2$ and $C_t = 0.9$). For the ADEPT method we selected ID E (DFF-smoothed 75^{th} quantile trace with $H_t = 0.2$ and $C_t = 0.98$). We selected these parameterizations due to their performance in both the baseline and pharmacologically manipulated conditions (Tables 3.6 and 3.7), as well as from visual inspection of Figures 3.8 and 3.9.

Figures 3.10 and 3.11 display boxplots with points comparing number of events in ROIs in this study at each condition. Each panel contains one of these plots, with column facets for each method and row facets conditioning

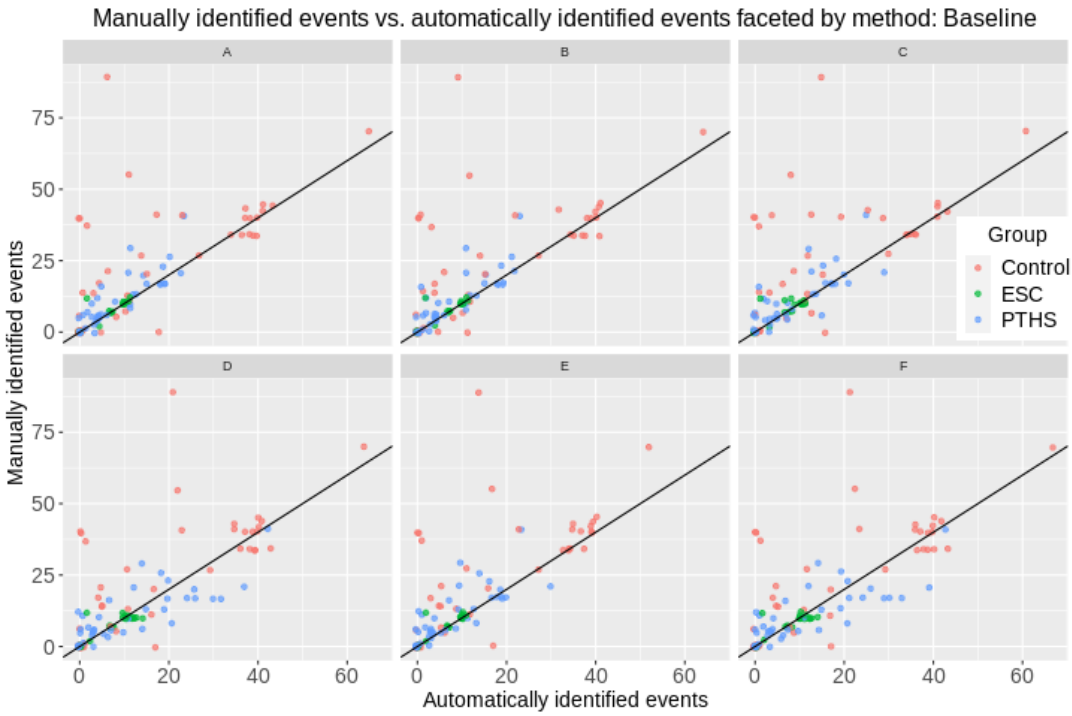


Figure 3.8: Manually identified events versus automatically identified events in the baseline condition faceted by method ID and colored by group. Each point represents a ROI identified from manual cell segmentation. y-axis: the number of events identified by manual peak detection. x-axis: the number of events identified by automated peak detection for the method and parameterization. The method and parameterization is identified by the letter ID in the top panel of each faceted plot. The method and parameterization ID correspond to each letter ID is described in Tables 3.6 and 3.7. In this figure, we visualize only the results from ROIs in the baseline condition. We see that all automated methods tended to underestimate the number of events of ROIs with a large number of manually identified events. Additionally, we see that most of these cases occurred in ROIs from healthy control donors. The black line in each plot is the identity line.

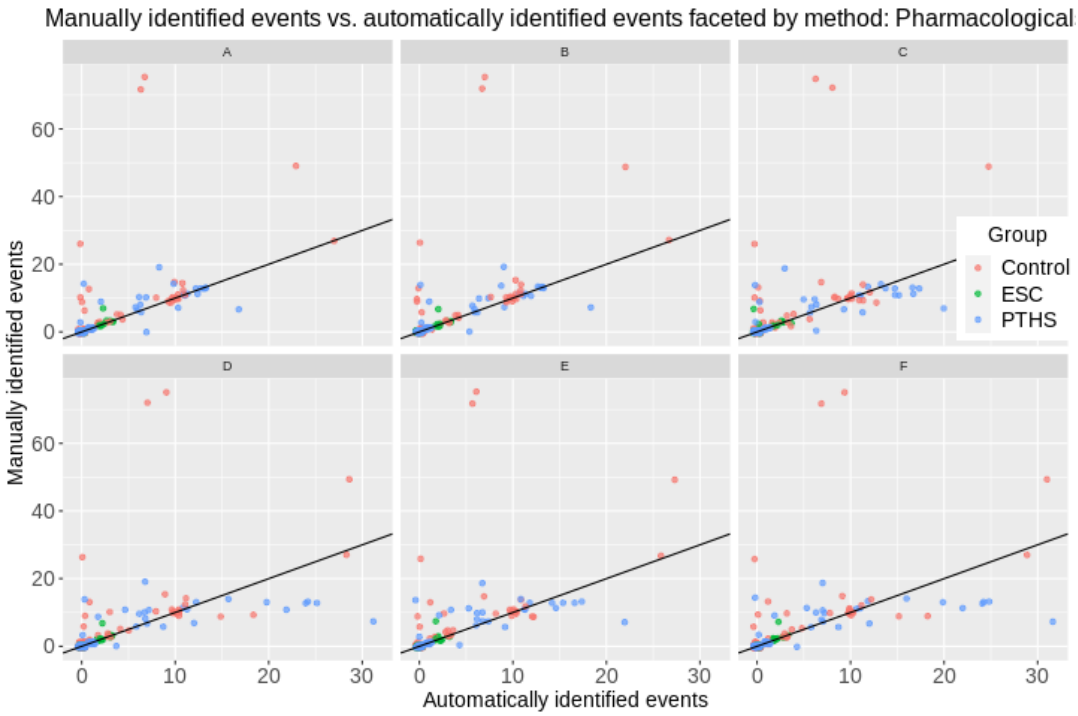


Figure 3.9: Manually identified events versus automatically identified events in pharmacologically induced conditions faceted by method ID and colored by group. Each point represents a ROI identified from manual cell segmentation. y-axis: the number of events identified by manual peak detection. x-axis: the number of events identified by automated peak detection for the method and parameterization. The method and parameterization is identified by the letter ID in the top panel of each faceted plot. The method and parameterization ID correspond to each letter ID is described in Tables 3.6 and 3.7. In this figure, we visualize only the results from ROIs in the pharmacologically-induced conditions. We see that all automated methods tended to underestimate the number of events of ROIs with a large number of manually identified events. Additionally, we see that most of these cases occurred in ROIs from healthy control donors. The black line in each plot is the identity line.

on activity. In Figure 3.10, we see these plots for ROIs identified in manual cell segmentation, and in Figure 3.11, we see these plots for ROIs identified by the median-eccentricity automated cell segmentation. We note that the far left column facet of both figures represents these boxplots for the manual event identification, which was only performed on ROIs identified in manual cell segmentation. The purpose of these plots is to see if automated methods produce similar distributions of events in each condition as the fully manual pipeline, and some aspects of the distributions are preserved. For instance, in each facet of both figures, in the baseline conditions, activity was more variable across ROIs and that we observed more highly active ROIs in the baseline condition, with reduced activity in each subsequent pharmacologically induced condition.

Figures 3.12, 3.13, and 3.14 display the events identified for various ROIs at the baseline condition. The figures show that with ideal trace plots, both automated methods perform very well with respect to identifying events and identifying them at approximately the correct time. However, in highly active ROIs (Figure 3.14), both methods can severely underestimate the number of events in a ROI over the time series. This result was concerning to the overall effectiveness of these methods in identifying activity events in highly active ROIs. The data in the study, however, were recorded at relatively slow temporal resolution (4 Hz). Data recorded at higher temporal resolutions may provide better data for identifying activity in highly-active ROIs.

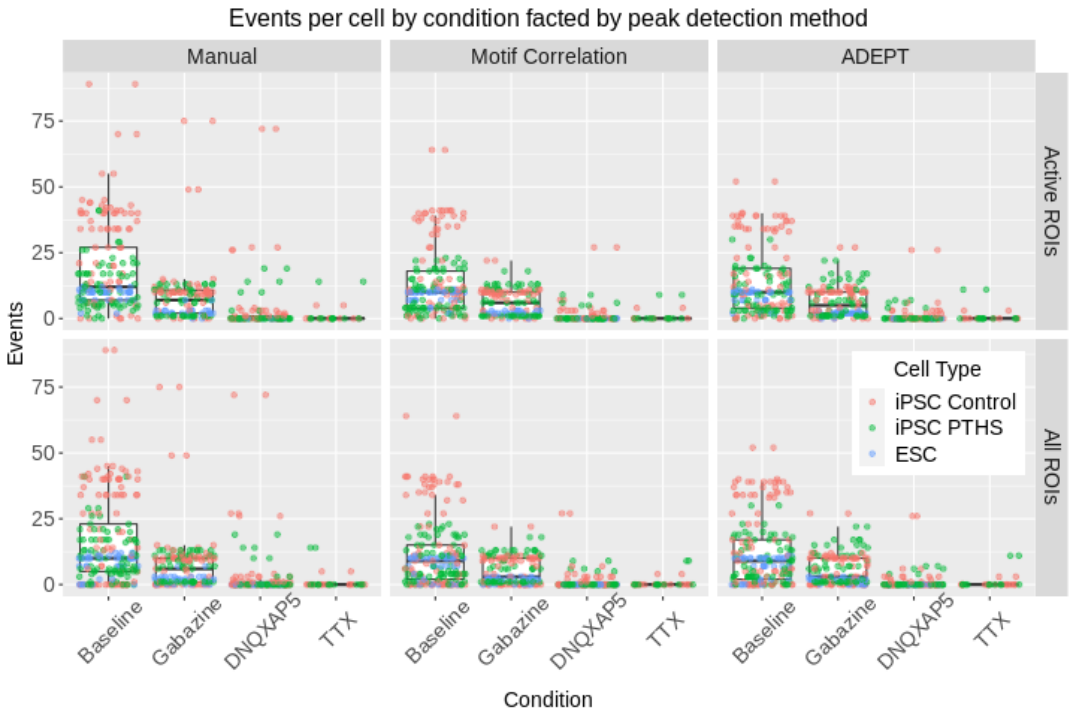


Figure 3.10: Boxplots of events per ROI versus condition using manual cell segmentation. Each point represents the number of events from single ROI in the study and are colored by donor group (neurotypical controls, PTHS, ESC). The plots are faceted by peak detection method and activity. The left column displays manually identified events. The middle column displays events identified by motif correlation. The right column displays events identified by ADEPT. The top row displays only ROIs that were identified as being "active", or having at least one event identified by the given method in any condition. The bottom row displays all ROIs. It should be noted that this means that, in the bottom row, each point in the manually peak detection plot has a corresponding point in each of the other plots. In the top row, this is only true of ROIs where at least one event was manually identified and at least one event was automatically identified for a given method.

Figure 3.11: Boxplots of events per ROI versus condition using median-eccentricity automated cell segmentation. Each point represents the number of events from single ROI in the study and are colored by donor group (neurotypical controls, PTHS, ESC). The plots are faceted by peak detection method and activity. The left column displays manually identified events. The middle column displays events identified by motif correlation. The right column displays events identified by ADEPT. The top row displays only ROIs that were identified as being "active", or having at least one event identified by the given method in the baseline condition. The bottom row displays all ROIs. Since manual event detection was only performed on the manually identified ROIs, the points in the plots in the left column come from different regions of interest than those in the middle and right columns.

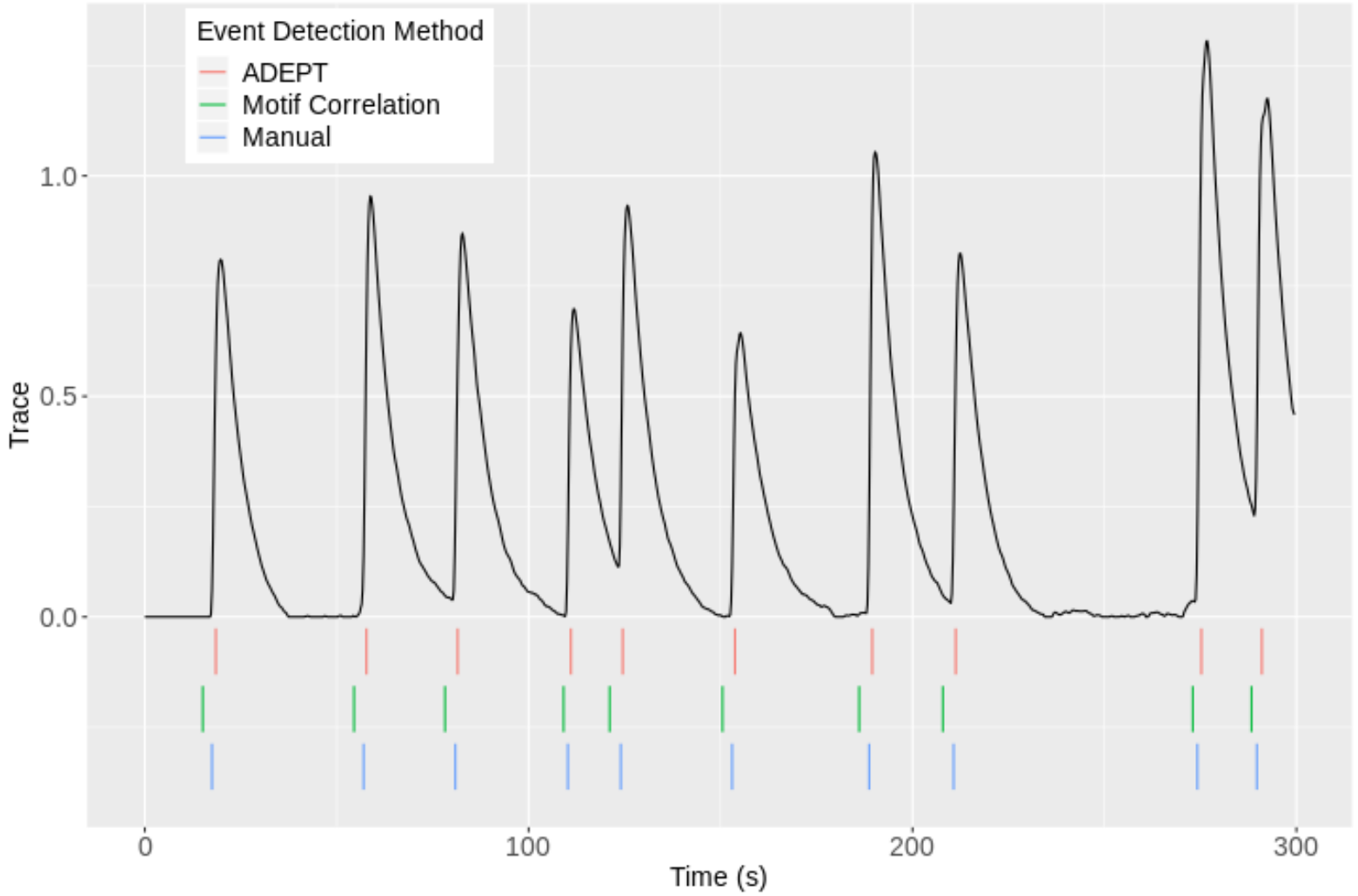


Figure 3.12: Event identification: ideal performance. DFF-smoothed trace plot of a ROI identified from an embryonic stem cell (ESC)-generated cell line at baseline with event identification from each method as tick marks on the bottom the plot. Each event detection method used a different trace summary statistic, however, for joint visualization we display the the DFF-smoothed mean trace on the y-axis. The x-axis displays the timepoint in seconds of the time series. We see this is an ideal case, where each automated method identifies an event at approximately the same time as each manually identified event.

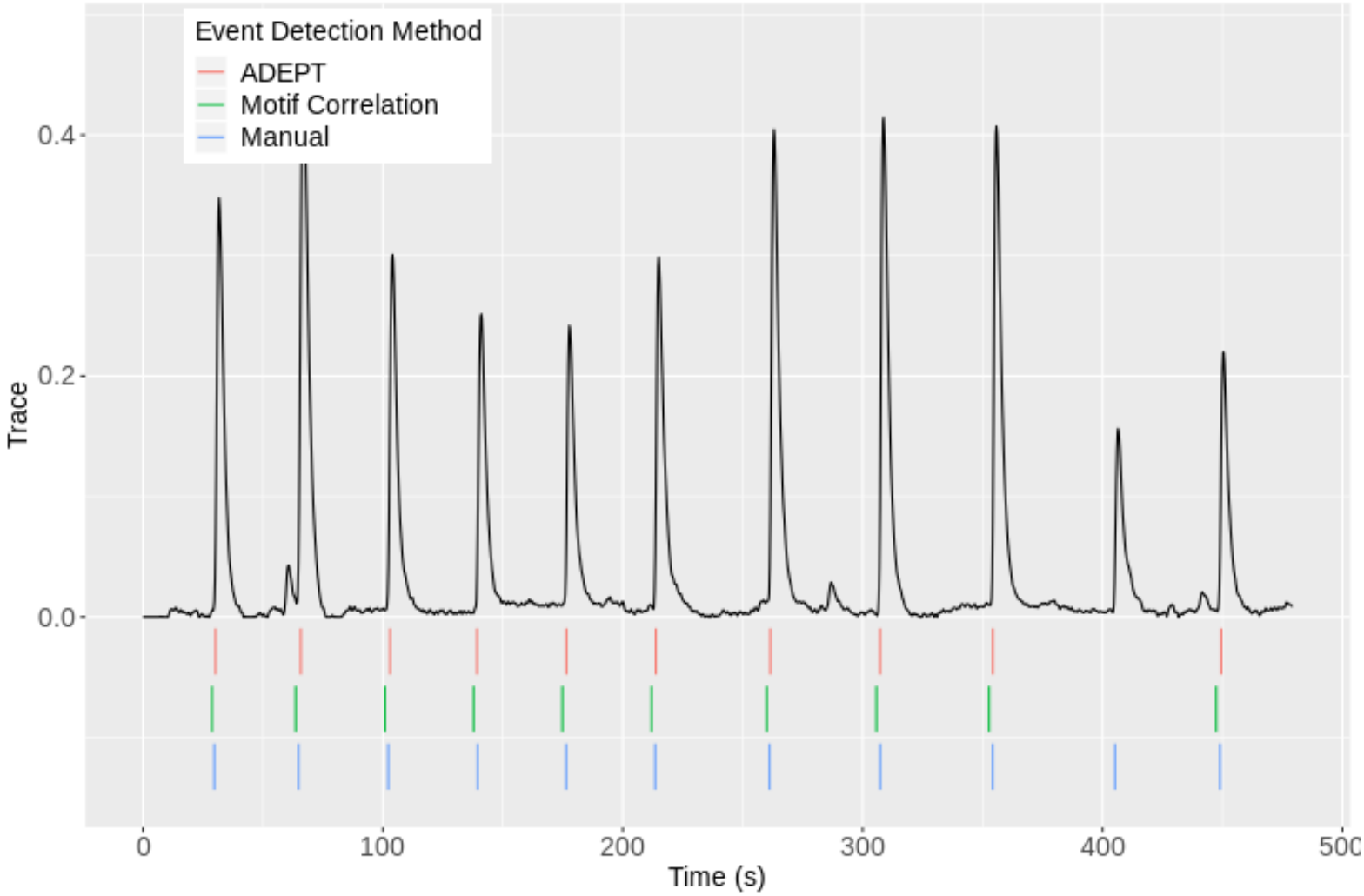


Figure 3.13: Event identification: moderate performance. DFF-smoothed trace plot of a ROI from an embryonic stem cell (ESC)-generated cell line at baseline with event identification from each method as tick marks on the bottom the plot. Each event detection method used a different trace summary statistic, however, for joint visualization we display the the DFF-smoothed mean trace on the y-axis. The x-axis displays the timepoint in seconds of the time series. We see, in this case both automated events miss one event just after the 400 second timepoint. This event was missed due to the height threshold parameter (H_t).

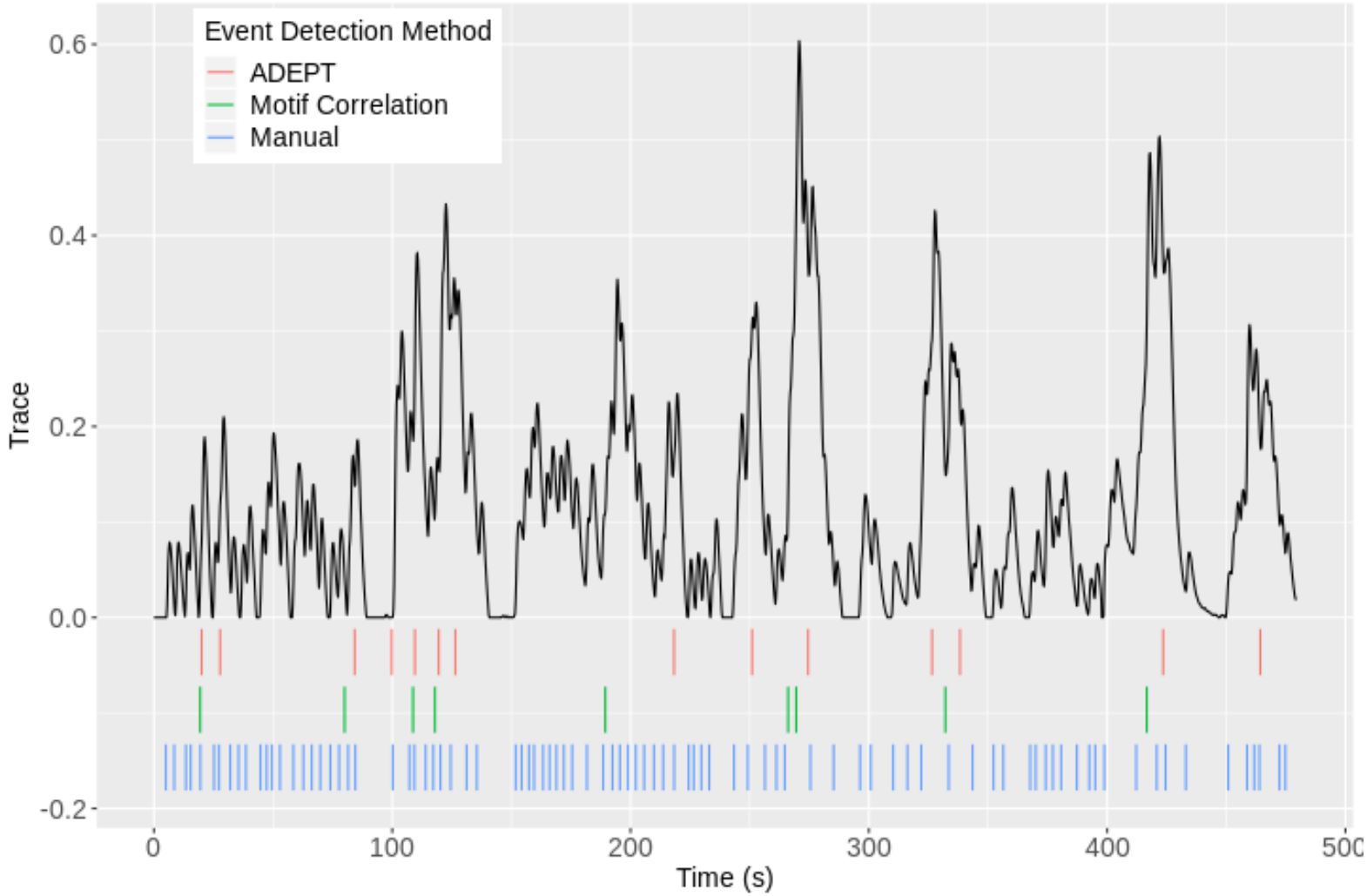


Figure 3.14: Event identification: poor performance with highly active ROIs. DFF-smoothed trace plot of ROI from a healthy control at baseline with event identification from each method as tick marks on the bottom the plot. Each event detection method used a different trace summary statistic, however, for joint visualization we display the the DFF-smoothed mean trace on the y-axis. The x-axis displays the timepoint in seconds of the time series. We see, in this case both automated events miss many of the events manually identified. This is a good representation of how events in highly active ROIs are often underestimated.

3.3.2 Statistical models: effect of pharmacological conditions estimated reasonably under optimized parameterizations

We assessed model performance by comparing downstream statistical inferences using to automated pipelines to results generated from the manual pipeline. First, we examined the results of regression model 2.2 when fit using activity data generated from the manual pipeline (Table 3.8). In the study, we observed that all pharamacological treatments were estimated to depress activity, and that within each pharmacological substance, these effects were estimated to be very similar whether when analyzing all cells/ROIs or only the subset of active cells. Overall, ROIs in the Gabazine condition were estimated to be 0.393 (95% CI: $[0.356, 0.435]$) times as active as ROIs in the baseline condition. We expected Gabazine to promote activity in this study as an antagonist of inhibitory GABA-a receptors. However, these results are not entirely surprising as there have been multiple findings of excitatory GABA-a receptors, particularly in immature or developing neuronal cells (Taketo and Yoshioka, 2000; Cherubini, Gaiarsa, and Ben-Ari, 1991). We further note that these cultured neurons have variable maturity even within the same collection of cells (Burke et al., 2020) and that globally these neurons resemble those from prenatal and not postnatal life (Hoffman et al., 2017). ROIs in the DNQXAP5 condition were estimated to be about 0.116 (95% CI: $[0.099, 0.135]$) times as active as ROIs in the baseline condition and ROIs in the TTX condition were estimated to be about 0.054 (95% CI: $[0.034, 0.086]$) times as active as ROIs in the baseline condition. Both of these results conformed to our expectations regarding directionality of effects.

		Estimates (95%CI)	
Effect	Cells	Coefficient (β_k)	Effect (e^{β_k})
Gabazine	All	-0.933 ([-1.034, -0.832])	0.393 ([0.356, 0.435])
Gabazine	Active	-0.932 ([-1.033, -0.831])	0.394 ([0.356, 0.435])
DNQXAP5	All	-2.158 ([-2.310, -2.005])	0.116 ([0.099, 0.135])
DNQXAP5	Active	-2.156 ([-2.309, -2.004])	0.116 ([0.099, 0.135])
TTX	All	-2.910 ([-3.368, -2.452])	0.054 ([0.034, 0.086])
TTX	Active	-2.910 ([-3.368, -2.452])	0.054 ([0.034, 0.086])

Table 3.8: Table of results of model 2.2 fit on activity data generated from the manual pipeline. We consider these estimates to be our best estimate of the truth. In the table we see the regression coefficient (β_k) for each pharmacological treatment from the model, as well as the corresponding estimate of the relative rate of activity for ROIs in each pharmacological treatment (e^{β_k}). We display 95% confidence intervals with each estimate. We display these estimates from models using all ROIs, as well as models using only ROIs that were identified as active (having at least one event across all conditions). In the table, we see that all pharamacological substances in this study were estimated to depress activity, and that within each pharmacological substance, these effects were estimated to be very similar whether we subset on active cells or not.

In this study, we considered these estimates to be the best estimates of the true effects of these pharmacological substances, and compared results from the same model fit on activity events identified from automated pipelines (Figure 3.15). Using the optimized parameterizations for each event detection method, we obtained reasonable estimates of the effect of each pharmacologically-induced condition on activity in ROIs when compared to estimates from manual event detection. The estimates from both automated event detection methods preserved the inference that all pharmacological substances employed in this study resulted in decreased activity in ROIs.

3.3.2.1 Automated event detection revealed differences between some of the pharmacological agents

Table 3.8 and Figure 3.15 both suggest that TTX had the strongest depressing effect of activity, followed by DNQXAP5 and Gabazine. We tested the differences in effects between pharmacologicals using the `linearHypothesis` function from the `car` package in `R` (Fox, 2015; Fox and Weisberg, 2019; Hand and Taylor, 1987; O'Brien and Kaiser, 1985) under model 2.2 for each method. Using manual event detection we observed strong evidence that TTX had a stronger effect of depression of activity than DNQXAP5 (all p-values < 0.01) and Gabazine (all p-values $< 1 \times 10^{-15}$). We also observed strong evidence that DNQXAP5 had a stronger effect of depression of activity than Gabazine (all p-values $< 1 \times 10^{-15}$). Using either automated event detection method, we observed strong evidence that TTX had a stronger effect of depression of activity than Gabazine (all p-values $< 1 \times 10^{-6}$), regardless of cell segmentation method. Using either automated event detection method, we also observed strong evidence that DNQXAP5 had a stronger effect of depression of activity than Gabazine (all p-values $< 1 \times 10^{-15}$), regardless of cell segmentation method. However, using either automated event detection method, we did not observe strong evidence of a difference in the effects of TTX and DNQXAP5 (all p-values > 0.1), regardless of cell segmentation method.

3.3.2.2 Estimates from motif correlation method were slightly closer to best estimates

We did observe some apparent bias in these estimates, as the automated peak detection methods appear to consistently estimate a lower relative activity

rate in ROIs under the DNQXAP5 condition than than results from manual event detection. In addition, the automated peak detection methods appear to consistently estimate a higher relative activity rate in ROIs under the Gabazine condition than than results from manual event detection. The absolute difference between the estimates from the optimized motif correlation method and manual event detection was consistently less than that between the ADEPT method and manual event detection (Table 3.9). We were able to formally test these differences by fitting model 2.2 using the data from all event detection methods and allowing an interaction term between event detection method and the effects of each pharmacological agent (β_k, $k \in \{1, 2, 3\}$). After fitting this model, we tested the difference between the automated methods' estimate of pharmacological effect using the `linearHypothesis` function. In this test, we did not observe strong evidence of a difference between results each of the automated methods relative to results from manual event detection (all p-values > 0.4).

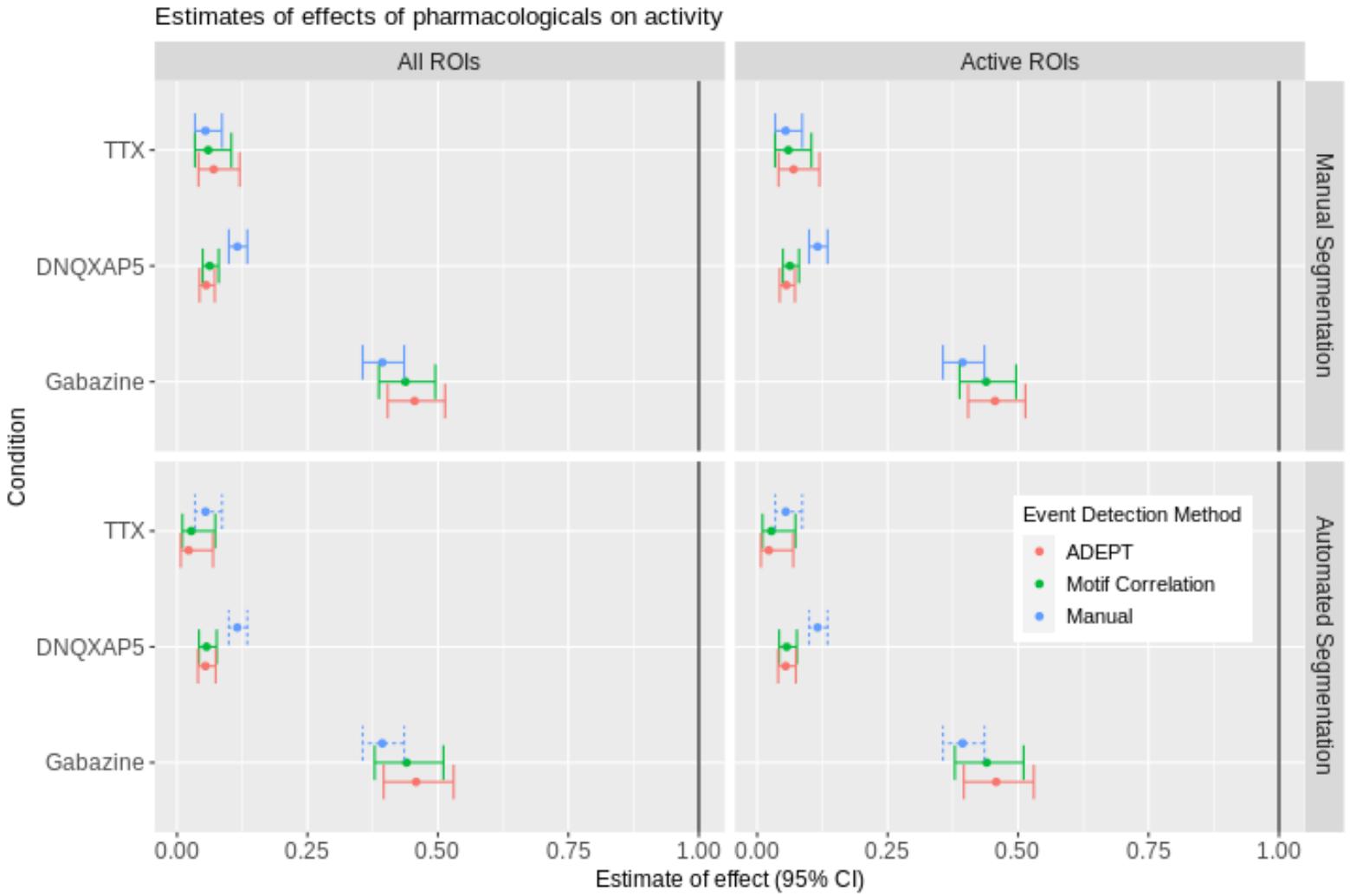


Figure 3.15: Comparison of inference from model 2.2 between results from manual event detection and automated event detection. In this figure we see the estimates of the relative activity rates of ROIs in each pharmacologically-induced condition compared to the baseline condition (e^{β_k}). The colors in this plot correspond to the event detection method. The points represent the points represent the point estimate of each effect from the model fit on data from each event detection method. The error bars represent the 95% confidence interval of each estimate. The left column displays results from the model fit on results from all ROIs and the right column displays results from the model fit on results from ROIs identified as being active. The top row displays results from the model fit on ROIs from manual cell segmentation and the bottom row displays results from the model fit on ROIs from automated cell segmentation using the median-eccentricity method. Since manual event detection was only performed on ROIs from manual cell segmentation, we display dashed lines for the results from manual event detection in the bottom row. We see in this figure that both methods, with the optimal parameterizations chosen in this study, give estimates of the effects of interest that are reasonably close to those drawn from results of manual event detection. Furthermore, under both methods, the inference that all pharmacological substances employed in this study reduced activity rate in the ROIs (3.8) is preserved.

Method	Estimate of Bias of Coefficient (β_k)		
	Gabazine	DNQXAP5	TTX
All ROIs with Manual Cell Segmentation			
Motif Correlation	**0.106**	**-0.613**	**0.092**
ADEPT	0.146	-0.728	0.254
All ROIs with Automated Cell Segmentation			
Motif Correlation	**0.111**	**-0.717**	**-0.692**
ADEPT	0.152	-0.754	-0.901
Active ROIs with Manual Cell Segmentation			
Motif Correlation	**0.109**	**-0.611**	**0.090**
ADEPT	0.147	-0.728	0.249
Active ROIs with Automated Cell Segmentation			
Motif Correlation	**0.111**	**-0.717**	**-0.692**
ADEPT	0.152	-0.754	-0.901

Table 3.9: Table of estimates of bias of coefficients from model 2.2 for each automated event detection method. Let β_k represent the model's coefficient estimates from the results of manual event detection, as we considered these to be the best estimate of the truth. Let $\hat{\beta}_{k,m}$ be these results from a given automated event detection method m. We estimate the bias of the coefficient estimates from method m to be $\hat{\beta}_{k,m} - \beta_k$. In this table we see that for every pharmacological condition, the motif correlation method consistently produced an estimate of β_k that was closer to the estimate from manual event detection than the ADEPT method, as denoted by the bolded values.

Chapter 4

Discussion and Conclusion

Determining the number of events from a set of cells in calcium imaging data is an exciting and challenging problem with several important processing steps. Quantifying activity in a large group of cells in a single experiment allows us to draw inference on the difference in activity rate between groups or pharmacologically-induced conditions.

In this study, we identified cells using a red channel image representing a measurement of the expression of the mRuby protein. We examined multiple cell segmentation methods and proposed a method that identifies cells as ROIs by thresholding by image standard deviations above the median of the red channel image. We also proposed an eccentricity-based cell refinement method for eliminating potential ROIs that we do not believe to be cells. We proposed this method because it performs preferably or similarly to other methods examined (Table 3.1), with a much faster processing time (Table 3.3). We recommend using a large standard-deviation based threshold in this method (between 6 and 10), as performance relative to manual cell segmentation results appeared to be stable within this range. However, when selecting

a threshold, we recommend some inspection of fluorescence data from the reference image used to identify cells. Some calcium imaging experiments may use fluorescence data from expression of a different protein than the one used in this study, or different microscope settings (i.e. spatial resolution), for identifying ROIs. These differences must be considered when selecting a proper threshold for identifying ROIs.

Additionally, we proposed a method for motion correction or drift. In this method, we started by estimating the translation matrix of each frame in a time series to the previous frame, then composed these matrices to register all frames of the time series to the space of the first frame. Finally, we estimated the translation matrix of the registered time series to the red channel image and register the time series to the space of the red channel image. In Figure 3.6, we see that this method is reasonable, even in time series recordings where significant drift occurred.

We also examined two primary methods for identifying events from trace statistics extracted from a time series recording of calcium imaging data. For each method, we employed a variety of trace statistics and parameterizations. First, we evaluated these methods based on their agreement with manual event detection in the number of events identified for a given ROI in a given cell in a recording. We optimized each method, by choosing the trace statistics and parameterizations that most agreed with manual event detection. Finally, estimated the effect on neuronal activity of three pharmacologically-induced conditions using events identified from both optimized automated event detection methods using a generalized linear mixed-effects regression model

(model 2.2). We continued to evaluate each event detection method by the similarity of these estimates and statistical inference drawn to those from like models fit on results from a manual pipeline. We saw that, using optimized parameterizations, both the motif correlation and ADEPT event detection method provided estimates and inferences of these effects that were reasonable when compared to the estimates and inferences drawn from a manual pipeline (Figure 3.15).

We see, however, an influence of parameter selection on downstream statistical inference. In this study, we employed 2,400 unique combinations of trace statistics and parameters for each event detection method (Figure 2.3). We also imposed tolerances on each method and parameterization's tendency to falsely identify inactive ROIs as active or falsely identify active ROIs as inactive. In order to fall within reasonable tolerances, we recommend using trace statistics that have been smoothed using DFF and employing global static height threshold (Figure 3.4). Even after focusing on parameterizations that used DFF and a global static height threshold, we still employed 400 unique parameterizations to each event detection method. In future work we would like to explore how stable the estimates and statistical inferences of the effects of these pharmacological agents were across these 400 parameterizations. We started this by visualizing the densities of the estimates of the effects (Figure 4.1) and the z-statistics generated from `glmer` (Figure 4.2) when fitting model 2.2 on event results from each method using parameterizations employed in this study. Figure 4.1 is meant to display the stability of the estimate of each pharmacological agent's effect from results of each method across

various parameterizations. We see in this figure that both methods tend to underestimate Gabazine's effect of depressing ROI activity in this study, and overestimate DNQXAP5's effect of depressing ROI activity. The distributions appear to be similar for each event detection method when employed on ROIs from manual cell segmentation, indicating that in this case the two methods had similar stability in their estimates across parameterizations. However, we do see some differences in the distributions of the estimates between the methods when employed on ROIs from automated cell segmentation, suggesting there could be important differences between the methods when employed in a fully automated pipeline. Figure 4.2 is meant to display the stability of the inference of each pharmacological agent's effect from results of each method across various parameterizations. In this figure, we see that both methods tend to underestimate the statistical significance of the evidence of the effect of every pharmacological in this study. We also see that the optimized parameterizations we chose in this study tended to be on the left side of the distribution of z-statistics. This suggests that we chose parameters that more closely aligned with our best estimates of the truth than most of the parameterizations employed in this study. We do see in both figures that the effects of all pharmacological agents were estimated to represent decreasing activity in the vast majority of parameterizations, conforming with our best estimates of the truth. This supports that these methods are appropriate for identifying relative ROI activity from calcium imaging data across these pharmacological conditions. Despite this result, the results still suggest that selecting proper parameterizations was crucial to generating estimates and inference that reasonably approximated our best estimates of the truth from

manual event detection. In future work, we will investigate what parameters most explain the variance in the estimates and statistical inferences drawn from results of each method.

One advantage of the ADEPT method is that it allows for parameterization to be applied after processing. However, for calcium imaging data similar to the data recorded in this study, we recommend using the motif correlation method for event detection with the parameters chosen in this study (DFF-smoothed median trace with correlation threshold $C_t = 0.9$ and static height threshold $H_t = 0.2$), as models using results from this method more often agreed with the statistical inferences drawn from models fit on results from a manual pipeline (Table 3.9). However, microscope settings (i.e. spatial and temporal resolution) may influence the effectiveness of the methods and parameterizations. We suggest some manual inspection of trace plots generated from calcium imaging data when selecting trace statistics, C_t and H_t. Future work includes applying the methods in this study to data sets from different subjects and species, such as calcium imaging of *in vivo* mouse cells. Additionally, we may want to examine the effectiveness of these methods on calcium imaging data from different microscopes with different spatial and temporal resolutions.

Overall, we evaluated a series of methods and parameter choices to improve automated calcium imaging data analysis, and established best practices for implementing these techniques on these data. We recommend using an automated cell segmentation using median z-scored images with a cell refinement based on shape eccentricity, though this method may miss a number

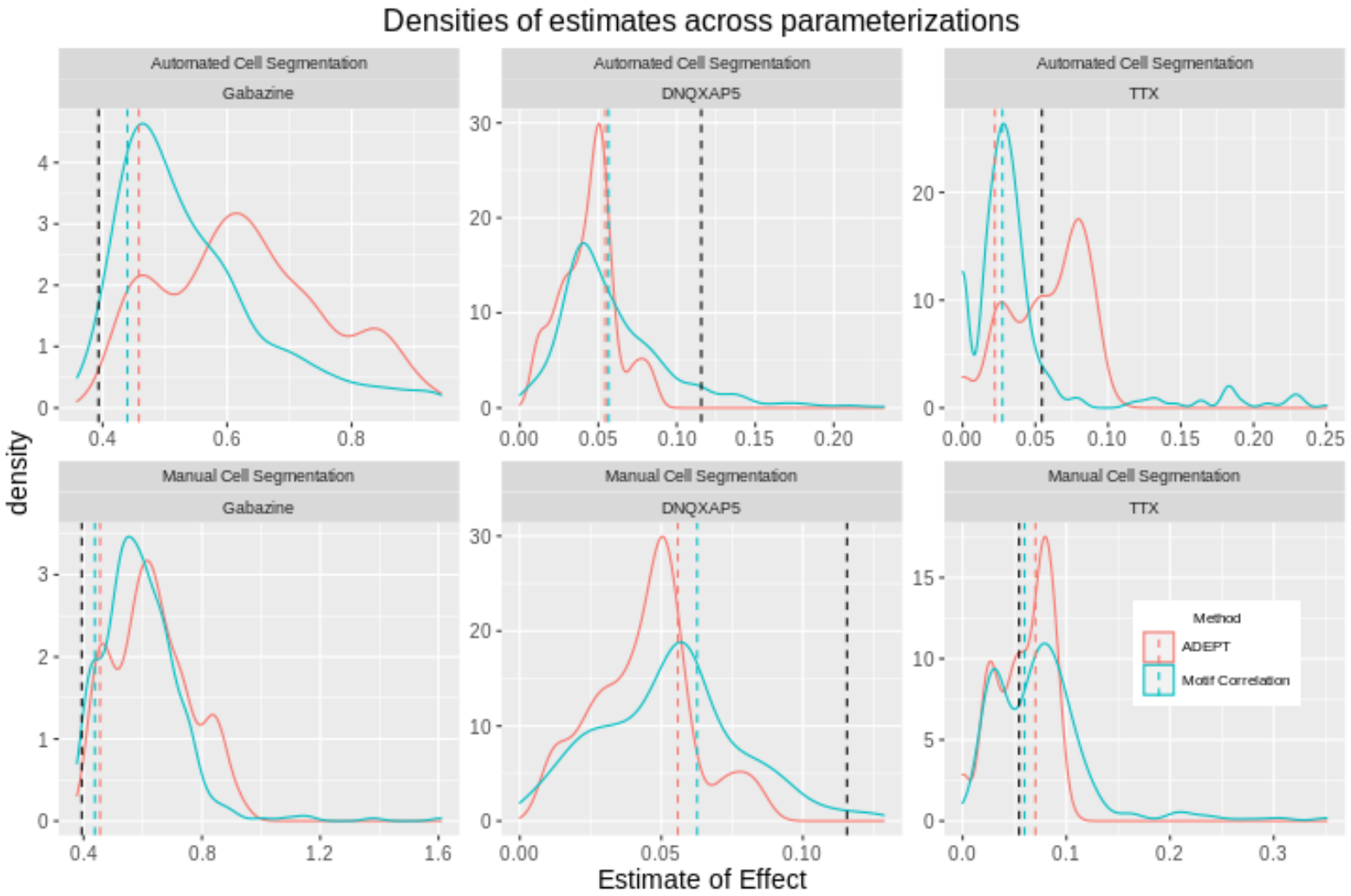


Figure 4.1: Densities of of the estimate effect of each pharmacological on activity in ROIs across 400 parameterizations for each method. For this figure, we fit model 2.2 on activity event results from each model with all 400 parameterizations that employed DFF smoothing and a global static height threshold. We present the estimate of the effect of each pharmacological on ROI activity (e^{β_k}). This figure is meant to display the stability of the estimate of each pharmacological agent's effect generated from models across parameterizations. In the left column, we see estimates for the effect of Gabazine, in the middle column estimates of the effect of DNQXAP5, and in the right column estimates of the effect of TTX. In the bottom row, we see estimates from ROIs selected in manual cell segmentation, and in the top row estimates from ROIs selected in automated cell segmentation using the median-eccentricity method. The colored dashed lines represent the estimate generated from the optimized parameterizations employed for each method in this study (Figure 3.15). The black dashed line represents the estimate generated from manual event detection.

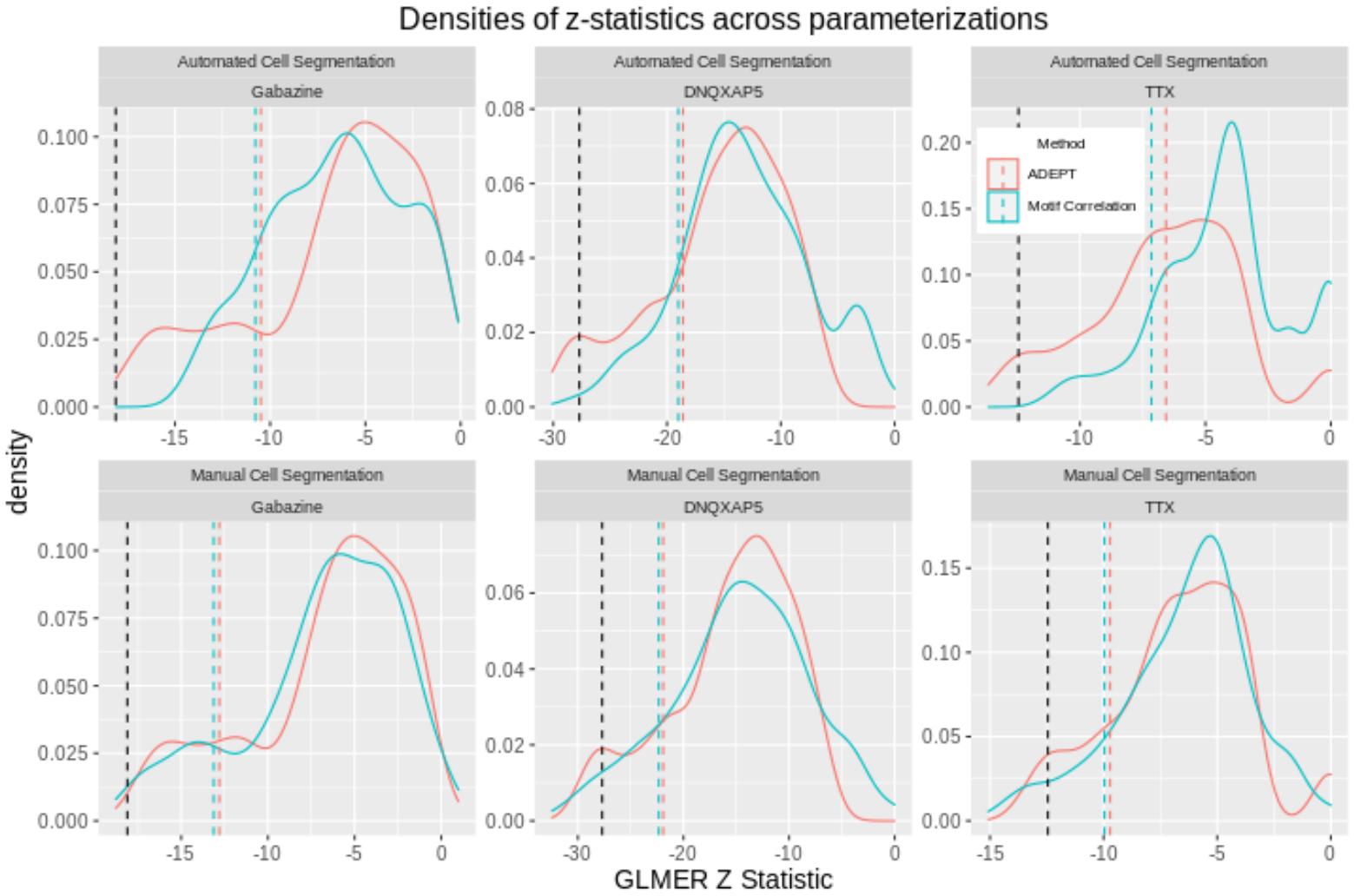


Figure 4.2: Densities of of the z-statistics generated from model 2.2 400 parameterizations for each method. For this figure, we fit model 2.2 on activity event results from each model with all 400 parameterizations that employed DFF smoothing and a global static height threshold. We present the z-statistics generated from `glmer` for the effect each pharmacological on ROI activity. This figure is meant to display the stability of the inference of each pharmacological agent's effect generated from models across parameterizations. In the left column, we see statistics for the effect of Gabazine, in the middle column statistics of the effect of DNQXAP5, and in the right column statistics of the effect of TTX. In the bottom row, we see estimates from ROIs selected in manual cell segmentation, and in the top row estimates from ROIs selected in automated cell segmentation using the median-eccentricity method. The colored dashed lines represent the estimate generated from the optimized parameterizations employed for each method in this study (Figure 3.15). The black dashed line represents the estimate generated from manual event detection. It should be noted that there were a very small amount of extreme negative outliers in this result, and so only z-statistics greater than or equal to -100 are displayed in this figure for visualization purposes.

of relevant cells, so it should not be used in cell counting. We recommend combining this with a motif-based correlation detection methods for peaks and events, though adding additional motifs may be useful for different experiments.

Bibliography

Grienberger, Christine and Arthur Konnerth (2012). "Imaging calcium in neurons". In: *Neuron* 73.5, pp. 862–885.

Ahrens, Misha B, Michael B Orger, Drew N Robson, Jennifer M Li, and Philipp J Keller (2013). "Whole-brain functional imaging at cellular resolution using light-sheet microscopy". In: *Nature methods* 10.5, p. 413.

Dana, Hod, Boaz Mohar, Yi Sun, Sujatha Narayan, Andrew Gordus, Jeremy P Hasseman, Getahun Tsegaye, Graham T Holt, Amy Hu, Deepika Walpita, et al. (2016). "Sensitive red protein calcium indicators for imaging neural activity". In: *Elife* 5, e12727.

Maher, Brady J and Joseph J LoTurco (2012). "Disrupted-in-schizophrenia (DISC1) functions presynaptically at glutamatergic synapses". In: *PLoS One* 7.3.

Mertens, Jerome, Qiu-Wen Wang, Yongsung Kim, X Yu Diana, Son Pham, Bo Yang, Yi Zheng, Kenneth E Diffenderfer, Jian Zhang, Sheila Soltani, et al. (2015). "Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder". In: *Nature* 527.7576, pp. 95–99.

Belinsky, Glenn S, Matthew T Rich, Carissa L Sirois, Shaina M Short, Erika Pedrosa, Herbert M Lachman, and Srdjan D Antic (2014). "Patch-clamp recordings and calcium imaging followed by single-cell PCR reveal the developmental profile of 13 genes in iPSC-derived human neurons". In: *Stem cell research* 12.1, pp. 101–118.

Birey, Fikri, Jimena Andersen, Christopher D Makinson, Saiful Islam, Wu Wei, Nina Huber, H Christina Fan, Kimberly R Cordes Metzler, Georgia Panagiotakos, Nicholas Thom, et al. (2017). "Assembly of functionally integrated human forebrain spheroids". In: *Nature* 545.7652, pp. 54–59.

Sweetser, David A, Ibrahim Elsharkawi, Lael Yonker, Marcie Steeves, Kimberly Parkin, and Ronald Thibert (2018). "Pitt-Hopkins Syndrome". In: *GeneReviews®[Internet]*. University of Washington, Seattle.

Amiel, Jeanne, Marlène Rio, Loïc De Pontual, Richard Redon, Valérie Malan, Nathalie Boddaert, Perrine Plouin, Nigel P Carter, Stanislas Lyonnet, Arnold Munnich, et al. (2007). "Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction". In: *The American Journal of Human Genetics* 80.5, pp. 988–993.

Brockschmidt, Antje, Unda Todt, Soojin Ryu, Alexander Hoischen, Christina Landwehr, Stefanie Birnbaum, Wilhelm Frenck, Bernhard Radlwimmer, Peter Lichter, Hartmut Engels, et al. (2007). "Severe mental retardation with breathing abnormalities (Pitt–Hopkins syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4". In: *Human molecular genetics* 16.12, pp. 1488–1494.

Forrest, Marc, Ria M Chapman, A Michelle Doyle, Caroline L Tinsley, Adrian Waite, and Derek J Blake (2012). "Functional analysis of TCF4 missense mutations that cause Pitt–Hopkins syndrome". In: *Human mutation* 33.12, pp. 1676–1686.

Sweatt, J David (2013). "Pitt–Hopkins syndrome: intellectual disability due to loss of TCF4-regulated gene transcription". In: *Experimental & molecular medicine* 45.5, e21–e21.

Van Balkom, Ingrid DC, Pieter Jelle Vuijk, Marijke Franssens, Hans W Hoek, and RAOUL CM HENNEKAM (2012). "Development, cognition, and behaviour in Pitt–Hopkins syndrome". In: *Developmental Medicine & Child Neurology* 54.10, pp. 925–931.

Zweier, Christiane, Heinrich Sticht, Emilia K Bijlsma, Jill Clayton-Smith, Susanne E Boonen, Alan Fryer, Marie T Greally, Lisbeth Hoffmann, Nicolette S den Hollander, Marjolijn Jongmans, et al. (2008). "Further delineation of Pitt–Hopkins syndrome: phenotypic and genotypic description of 16 novel patients". In: *Journal of medical genetics* 45.11, pp. 738–744.

Giovannucci, Andrea, Johannes Friedrich, Pat Gunn, Jeremie Kalfon, Brandon L Brown, Sue Ann Koay, Jiannis Taxidis, Farzaneh Najafi, Jeffrey L Gauthier, Pengcheng Zhou, et al. (2019). "CaImAn an open source tool for scalable calcium imaging data analysis". In: *Elife* 8, e38173.

Patel, Tapan P, Karen Man, Bonnie L Firestein, and David F Meaney (2015). "Automated quantification of neuronal networks and single-cell calcium dynamics using calcium imaging". In: *Journal of neuroscience methods* 243, pp. 26–38.

Jewell, Sean, Toby Dylan Hocking, Paul Fearnhead, and Daniela Witten (2018). "Fast nonconvex deconvolution of calcium imaging data". In: *arXiv preprint arXiv:1802.07380*.

Wen, Zhexing, Ha Nam Nguyen, Ziyuan Guo, Matthew A Lalli, Xinyuan Wang, Yijing Su, Nam-Shik Kim, Ki-Jun Yoon, Jaehoon Shin, Ce Zhang, et al. (2014). "Synaptic dysregulation in a human iPS cell model of mental disorders". In: *Nature* 515.7527, pp. 414–418.

Taketo, M and T Yoshioka (2000). "Developmental change of GABAA receptor-mediated current in rat hippocampus". In: *Neuroscience* 96.3, pp. 507–514.

Cherubini, Enrico, Jean L Gaiarsa, and Yehezkel Ben-Ari (1991). "GABA: an excitatory transmitter in early postnatal life". In: *Trends in neurosciences* 14.12, pp. 515–519.

Linkert, Melissa, Curtis T Rueden, Chris Allan, Jean-Marie Burel, Will Moore, Andrew Patterson, Brian Loranger, Josh Moore, Carlos Neves, Donald MacDonald, et al. (2010). "Metadata matters: access to image data in the real world". In: *Journal of Cell Biology* 189.5, pp. 777–782.

Kredel, Simone, Franz Oswald, Karin Nienhaus, Karen Deuschle, Carlheinz Röcker, Michael Wolff, Ralf Heilker, G Ulrich Nienhaus, and Jörg Wiedenmann (2009). "mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures". In: *PloS one* 4.2.

Haralick, Robert M and Linda G Shapiro (1992). *Computer and robot vision*. Vol. 1. Addison-wesley Reading.

Atherton, Tim J and Darren J Kerbyson (1999). "Size invariant circle detection". In: *Image and Vision computing* 17.11, pp. 795–803.

Yuen, HK, John Princen, John Illingworth, and Josef Kittler (1990). "Comparative study of Hough transform methods for circle finding." In: *Image and vision computing* 8.1, pp. 71–77.

Davies, ER (2005). *Machine Vision: Theory, Algorithms, Practicalities, 3rd*.

Meyer, Fernand (1994). "Topographic distance and watershed lines". In: *Signal processing* 38.1, pp. 113–125.

Soille, Pierre (2013). *Morphological image analysis: principles and applications*. Springer Science & Business Media.

Mattes, David, David R Haynor, Hubert Vesselle, Thomas K Lewellen, and William Eubank (2003). "PET-CT image registration in the chest using free-form deformations". In: *IEEE transactions on medical imaging* 22.1, pp. 120–128.

Jia, Hongbo, Nathalie L Rochefort, Xiaowei Chen, and Arthur Konnerth (2011). "In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons". In: *Nature protocols* 6.1, p. 28.

Bai, Jiawei, Jeff Goldsmith, Brian Caffo, Thomas A Glass, and Ciprian M Crainiceanu (2012). "Movelets: A dictionary of movement". In: *Electronic journal of statistics* 6, p. 559.

He, Bing, Jiawei Bai, Vadim V Zipunnikov, Annemarie Koster, Paolo Caserotti, Brittney Lange-Maia, Nancy W Glynn, Tamara B Harris, and Ciprian M Crainiceanu (2014). "Predicting human movement with multiple accelerometers using movelets". In: *Medicine and science in sports and exercise* 46.9, p. 1859.

Lin, Rake& Agrawal King-lp and Harpreet S Sawhney Kyuseok Shim (1995). "Fast similarity search in the presence of noise, scaling, and translation in time-series databases". In: *Proceeding of the 21th International Conference on Very Large Data Bases*. Citeseer, pp. 490–501.

Chan, Kin-Pong and Ada Wai-Chee Fu (1999). "Efficient time series matching by wavelets". In: *Proceedings 15th International Conference on Data Engineering (Cat. No. 99CB36337)*. IEEE, pp. 126–133.

Ge, Xianping and Padhraic Smyth (2000). "Deformable Markov model templates for time-series pattern matching". In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 81–90.

Akima, Hiroshi (1970). "A new method of interpolation and smooth curve fitting based on local procedures". In: *Journal of the ACM (JACM)* 17.4, pp. 589–602.

Akima, Hiroshi (1974). "A method of bivariate interpolation and smooth surface fitting based on local procedures". In: *Communications of the ACM* 17.1, pp. 18–20.

Karas, Marta, Marcin Straczkiewicz, William Fadel, Jaroslaw Harezlak, Ciprian M Crainiceanu, and Jacek K Urbanek (2019). "Adaptive empirical pattern transformation (ADEPT) with application to walking stride segmentation". In: *Biostatistics*. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxz033. eprint: https://academic.oup.com/biostatistics/advance-article-pdf/doi/10.1093/biostatistics/kxz033/30043987/kxz033.pdf. URL: https://doi.org/10.1093/biostatistics/kxz033.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Dice, Lee R (1945). "Measures of the amount of ecologic association between species". In: *Ecology* 26.3, pp. 297–302.

Burke, Emily E, Joshua G Chenoweth, Joo Heon Shin, Leonardo Collado-Torres, Suel-Kee Kim, Nicola Micali, Yanhong Wang, Carlo Colantuoni, Richard E Straub, Daniel J Hoeppner, et al. (2020). "Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs". In: *Nature communications* 11.1, pp. 1–14.

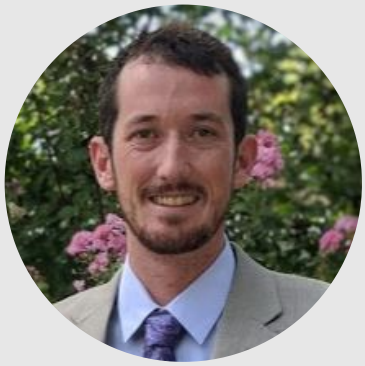
Hoffman, Gabriel E, Brigham J Hartley, Erin Flaherty, Ian Ladran, Peter Gochman, Douglas M Ruderfer, Eli A Stahl, Judith Rapoport, Pamela Sklar, and Kristen J Brennand (2017). "Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains". In: *Nature communications* 8.1, pp. 1–15.

Fox, John (2015). *Applied regression analysis and generalized linear models*. Sage Publications.

Fox, John and Sanford Weisberg (2019). *An R Companion to Applied Regression (Third)*.

Hand, David J and Colin C Taylor (1987). *Multivariate analysis of variance and repeated measures: a practical approach for behavioural scientists*. Vol. 5. CRC press.

O'Brien, Ralph G and Mary K Kaiser (1985). "MANOVA method for analyzing repeated measures designs: an extensive primer." In: *Psychological bulletin* 97.2, p. 316.



Joseph L. Catallini II

Data Scientist/Analyst

1910 Eastern Ave #1
Baltimore, MD 21231

(520) 306-7584

j.l.cat.6@gmail.com

/in/joseph-catallini-ii

joecat6

About Me

Data Analyst with 6+ years of professional experience pulling data from a variety of disparate sources and converting it into easily interpretable formats to explain trends and provide analyses and recommendations to optimize systems and processes. Experience includes contributing to a software development environment with a variety of languages and tools. Programming strengths include data extraction, analysis, and visualization, tool and GUI development, and process automation. Interested in applying advanced methods to cutting-edge research in biomedical technologies.

Skills

Programming Languages -

Python, R, MATLAB, SQL, C#, C++, Stan, Visual Basic, Java

OS & Tools

Windows, Linux, VxWorks, R Markdown, Jupyter Notebooks, Machine Learning, Image Processing, XML, XSLT, Microsoft Office, JIRA, Git, ClearCase, Agile, LaTeX, Reproducible Research

Foreign Languages

Spanish (Conversational)

Education

Johns Hopkins Bloomberg School of Public Health 2018-Present

ScM Candidate Biostatistics, Expected Graduation: May 2020

GPA: 3.94

Relevant Coursework: Methods in Biostatistics, Essentials of Probability and Statistical Inference, Statistical Computing, Introduction to Data Management, Bayesian Methods, Survival Analysis, Advanced Data Science

Research Interest: *Calcium imaging of neuronal cells*.

University of Arizona 2009-2013

BS Mathematics, Honors College, Cum Laude

GPA: 3.57

Minors: Spanish, Sports Management

Positions: Undergraduate Research Assistant, Math Tutor

Honors Capstone Thesis:*The Early Negro Leagues and Major League Baseball: A Comparative Analysis*, Advisor: Dr. Joseph Watkins

Experience

Johns Hopkins/Lieber Institute, Baltimore, MD Oct 2018-Present

Research Assistant

- Developing processing protocols and analysis methods of calcium imaging data measuring neuronal activity
- Producing data visualization tools to deliver to Research Scientists and Lab Technicians

Boeing, Saint Louis, MO Sep 2014-Aug 2018

Electrophysics Engineer/Scientist II

- Managed data- and testing-related projects and mentored junior-level engineers on testing procedures
- Automated build, installation, and testing procedures for software product development in an Agile environment
- Created data visualization tools and reports for software engineers and systems architects
- Leveraged data analysis and regression testing to optimize products and meet customer requirements

Ventana Medical Systems (Roche), Tucson, AZ Oct 2013-Sep 2014

Clinical Data Associate

- Performed data entry and quality control and developed dual-verification data entry tools
- Assisted pathologists and engineers in organizing studies, operating devices, and creating analysis tools to evaluate, troubleshoot, and optimize protocols for medical devices
- Developed algorithms for digital pathology software that utilized image processing and supervised learning to assess stain quality on tissue culture slides